The Degrees of Freedom Region of the MIMO Interference Channel with Shannon Feedback

Chinmay S. Vaze and Mahesh K. Varanasi

Abstract

The two-user multiple-input multiple-output (MIMO) fast-fading interference channel (IC) with an arbitrary number of antennas at each of the four terminals is studied under the settings of Shannon feedback, limited Shannon feedback, and output feedback, wherein all or certain channel matrices and outputs, or just the channel outputs, respectively, are available to the transmitters with a finite delay. While for most numbers of antennas at the four terminals, it is shown that the DoF regions with Shannon feedback and for the limited Shannon feedback settings considered here are identical, and equal to the DoF region with just delayed channel state information (CSIT), it is shown that this is not always the case. For a specific class of MIMO ICs characterized by a certain relationship between the numbers of antennas at the four nodes, the DoF regions with Shannon and the limited Shannon feedback settings, while again being identical, are strictly bigger than the DoF region with just delayed CSIT. To realize these DoF gains with Shannon or limited Shannon feedback, a new retrospective interference alignment scheme is developed wherein transmitter cooperation made possible by output feedback in addition to delayed CSIT is employed to effect a more efficient form of interference alignment than is feasible with previously known schemes that use just delayed CSIT. The DoF region for just output feedback, in which each transmitter has delayed knowledge of only the receivers' outputs, is also obtained for all but a class of MIMO ICs that satisfy one of two inequalities involving the numbers of antennas.

Index Terms

Degrees of freedom, delayed CSIT, feedback, interference channel, interference alignment, MIMO, Shannon feedback.

I. INTRODUCTION

THE characterization of the capacity of channels with feedback, where the channel outputs are known to the transmitter(s) with a finite delay, is a classical problem in information theory. For example, it is well known that feedback can not increase the capacity of a memoryless point-to-point channel [1]. Moreover, feedback can not increase the capacity of a point-to-point channel with additive, correlated Gaussian noise by more than one bit [1]. Interestingly, multi-user channels exhibit a different behavior. In particular, feedback can enhance the capacity of even the memoryless multiple access channel (MAC) [2], [3] but again this improvement is bounded for the Gaussian MAC [3]. There has also been much interest in characterizing the capacity region of other memoryless networks with feedback such as the broadcast channel (BC) [4]. However, due to the apparent intractability of such problems for more complex topologies, capacity *approximations* have been sought. Of these approximate capacity metrics, the degrees of freedom (DoF) region has received considerable attention. The DoF region denotes the rate of growth of the capacity region with respect to the logarithm of the signal-to-noise ratio (SNR) in the limit of asymptotically high SNR. For example, it can be deduced from [5] that for the 2-user Gaussian MIMO IC output feedback can not enhance the DoF region when there is perfect and instantaneous CSIT. In [6], the feedback capacity region is characterized to within a constant gap of 2 bits for the single-antenna (or SISO) IC. Further, it is well known that feedback fails to improve the DoF regions of the Gaussian MIMO MAC and the Gaussian MIMO BC. It is not yet known if there are networks with instantaneous CSIT for which (output) feedback enhances the DoF region.

While the DoF of networks under the idealized assumptions of perfect, often global, and instantaneous CSIT have been well studied for numerous networks, the much more conservative setting of isotropic fading with transmitters having no CSIT has recently been extensively studied as well [7]–[14]. Networks without CSIT but with (output) feedback have also been considered from which it is known that in the absence of CSIT feedback can enhance the DoF regions of the K-user BC [15]–[17], the 2-user SISO X channel [18], and 3-user SISO IC [10], [18]. Thus, unlike the instantaneous CSIT case, feedback can be beneficial even from the DoF perspective when there is no CSIT. This suggests that the benefit of feedback depends critically on the availability of CSIT since it is vastly different at the two extremes of having instantaneous CSIT and having no CSIT whatsoever.

Moving beyond models that are either too idealized on the one extreme, or too conservative on the other, we consider here the delayed CSIT model wherein the channel state varies independently across time and the transmitters know perfectly the *past* channel states (cf. [15], [16], [18]). For such a setting we investigate the question of whether output feedback can improve the DoF region. Clearly, this question can be definitively answered only for networks for which the DoF with (just) delayed CSIT are known, of which there are but few. Of all networks for which the DoF are known except for the MIMO IC, this question has so far been answered in the negative. In particular, it is known that with delayed CSIT output feedback can not increase (a) the sum-DoF of the K-user MISO BC with at least K transmit antennas [15], (b) the DoF region of the 2-user MIMO BC [16], (c) the sum-DoF of the 3-user MIMO BC with N antennas at all three receivers and at most $2N$ antennas at the transmitter [17] and (d) the $2 \times 2 \times 2$ interference network [19].

The only other exact characterization for the DoF region with (just) delayed CSIT is provided by the authors in [20] for the two-user MIMO IC with an arbitrary number of antennas at each of the four terminals. Consequently, by obtaining the complete DoF region of this general two-user MIMO IC with delayed CSIT and output feedback (i.e., under Shannon feedback), and showing that for some cases there is a strict enhancement of the DoF region over that with just delayed CSIT, we answer the question of whether output feedback can enhance the DoF region of a network with delayed CSIT for the first time in the affirmative.

In particular, it is shown here that if M_i and N_i, respectively, denote the number of antennas at the ith transmitter and the ith receiver of a two-user MIMO IC, then the DoF region with Shannon feedback is strictly bigger than the corresponding delayed-CSIT DoF region, if and only if one of the two inequalities, namely,

$$M_1 > N_1 + N_2 - M_2 > N_1 > N_2 > M_2 > N_2 \frac{N_2 - M_2}{N_1 - M_2}$$

or its symmetric counterpart (obtained by switching the user indices), holds. For MIMO ICs for which neither of these two inequalities holds output feedback does not improve the delayed CSIT DoF region.

To derive our main result, we first obtain an outer-bound to the DoF region with Shannon feedback. This outer bound is then shown to be achievable for all but the above described class of MIMO ICs using just delayed
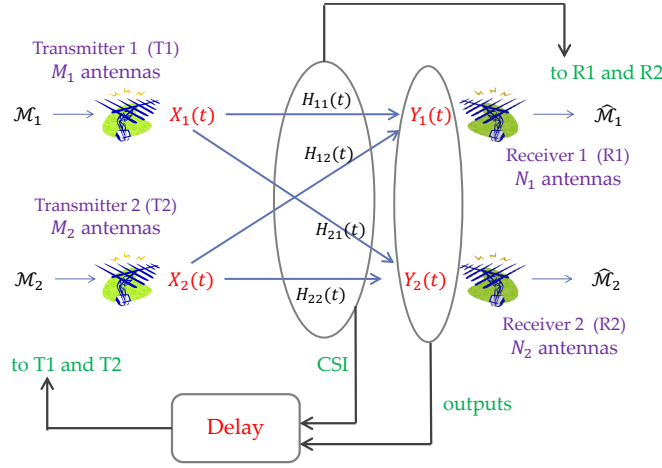
Fig. 1. The (M_1, M_2, N_1, N_2) MIMO IC with Shannon Feedback

CSIT. For the class where the DoF region is strictly larger than that with delayed CSIT, we develop a new retrospective interference alignment scheme in which each transmitter – using the side information available to it – reconstructs and transmits the previously transmitted signal of the other transmitter to provide an opportunity to its paired receiver to cancel the interference it encountered at a previous time instant, while simultaneously delivering new useful linear combinations to the unpaired receiver. Consequently, Shannon feedback induces a new form of transmitter cooperation which is key to realizing the DoF gains attainable with Shannon feedback over that with just delayed CSIT.

Moreover, it is seen that to achieve this more efficient interference alignment all of the channel matrices and outputs are not needed at both transmitters. In particular, two limited Shannon feedback settings are described that are sufficient to achieve the DoF region with Shannon feedback. It is also observed that if in addition to delayed CSIT the feedback is allowed to be some designable function of past channel outputs and states, a setting that is more optimistic than Shannon feedback, the DoF region doesn't expand over that of the DoF region in the Shannon feedback case. Furthermore, with just output feedback without any form of CSIT, it is shown that the DoF region is the same as that for delayed CSIT with the exception of a class of MIMO ICs characterized by one of two inequalities involving the numbers of antennas at the four terminals. For this class, the DoF region remains an open problem at this time.

The rest of the paper is organized as follows. Section II-A describes the model of MIMO IC under various assumptions about feedback, states the main results of this work on the DoF regions under these assumptions and provides an example of the new retrospective interference alignment scheme. Proofs of the results are contained in Sections III and IV and the appendix.

II. CHANNEL MODELS, MAIN RESULTS, AND IA WITH SHANNON FEEDBACK

In Section II-A, the MIMO IC model with Shannon feedback, limited Shannon feedback, output feedback and designable Shannon feedback are described. Section II-B contains the main results on the DoF regions under these settings. In Section II-C, we discuss how interference alignment can be achieved with Shannon and limited Shannon feedback. Section II-D provides some insight on the main results.

A. The MIMO IC with Shannon Feedback

The MIMO IC consists of two transmitters, denoted as T1 and T2, and their corresponding receiver, labeled R1 and R2, respectively, as in Fig. 1. The (M_1, M_2, N_1, N_2) MIMO IC is defined via the input-output relationships

$$Y_1(t) = H_{11}(t)X_1(t) + H_{12}(t)X_2(t) + W_1(t), \tag{1}$$
$$Y_2(t) = H_{21}(t)X_1(t) + H_{22}(t)X_2(t) + W_2(t), \tag{2}$$

where, at time t, $Y_i(t) \in \mathbb{C}^{N_i \times 1}$ is the signal received by the ith receiver; $X_i(t) \in \mathbb{C}^{M_i \times 1}$ is the signal transmitted by the ith transmitter; $H_{ij}(t) \in \mathbb{C}^{N_i \times M_j}$ is the channel matrix between the ith receiver and the jth transmitter;
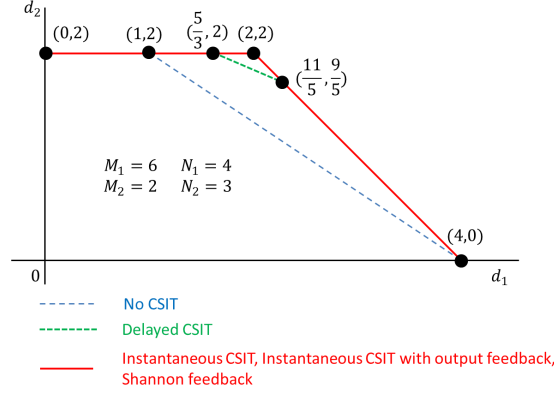
Fig. 2. Comparison of the DoF regions of a MIMO IC

$W_i(t) \in \mathbb{C}^{N_i \times 1}$ is the additive white Gaussian noise at the ith receiver; and there is a power constraint of P on the transmit signals, i.e., $\mathbb{E}||X_i(t)||^2 \leq P \ \forall \ i, t$.

For simplicity, we study here the case of additive white Gaussian noise, i.e., all entries of $W_i(t)$, $i = 1, 2$, are independent and identically distributed (i.i.d.) according the complex normal distribution with zero-mean and unit-variance (denoted, henceforth, as $\mathcal{CN}(0, 1)$). Further, we assume that the channel matrices are i.i.d. Rayleigh faded, i.e., all elements of all channel matrices are i.i.d. according to $\mathcal{CN}(0, 1)$ distribution (denoted as i.i.d. $\sim \mathcal{CN}(0, 1)$). Next, it is assumed that the realizations of additive noises and channel matrices are i.i.d. across time and that they are mutually independent of each other.

Throughout this paper, both receivers are taken to know all channel matrices perfectly. Since there is no delay constraint on decoding it is assumed, without loss of generality, that CSI is available instantaneously to the receivers. Further, all terminals are always assumed to know the distribution of the channel matrices.

We start by defining the term Shannon feedback (cf. [18]) and later consider other types of feedback. Here, the two transmitters are assumed to know the channel matrices and the channel outputs perfectly with a finite delay. This delay is taken to be of 1 symbol time without loss of generality. In particular, the channel matrices $\{H_{ij}(t)\}_{i,j=1}^{2}$ and the channel outputs $Y_1(t)$ and $Y_2(t)$ are taken to be known perfectly to both transmitters at time $t + 1$.

Notation: For $n \geq 0$, $\overline{H}(n) \stackrel{\triangle}{=} \{H_{11}(t), H_{12}(t), H_{21}(t), H_{22}(t)\}_{t=1}^{n}$ if $n \geq 1$ and $\overline{H}(n) = \phi$ ($\phi =$ some constant) if $n = 0$. Similarly, for each $i \in \{1, 2\}$, $\overline{Y}_i(n) = \{Y_i(t)\}_{i=1}^{n}$ if $n \geq 1$ and $\overline{Y}_i(n) = \phi$ if $n = 0$.

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two independent messages to be sent by T1 and T2, respectively, over a block length of n, where the message $\mathcal{M}_i$ is intended for the ith receiver. It is assumed that $\mathcal{M}_i$ is distributed uniformly over a set of cardinality $2^{nR_i(P)}$, $R_i(P) \geq 0$, when there is a power constraint of P at the transmitters. A coding scheme for blocklength n consists of two encoding functions $f_i^{(n)} = \{f_{i,t}^{(n)}\}_{t=1}^{n}$, $i = 1, 2$, such that

$$X_i(t) = f_{i,t}^{(n)}\left(\mathcal{M}_i, \overline{H}(t-1), \overline{Y}_1(t-1), \overline{Y}_2(t-1)\right) \text{ with } \mathbb{E}||X_i(t)||^2 \leq P \ \forall \ i, t,$$

and two decoding functions such that

$$\hat{\mathcal{M}}_i = g_i^{(n)}\left(\overline{Y}_i(n), \overline{H}(n)\right) \text{ where } i \in \{1, 2\}.$$

A rate tuple $(R_1(P), R_2(P))$ is said to be achievable if there exists a sequence (over n) of coding schemes such that probability of $\mathcal{M}_1 \neq \hat{\mathcal{M}}_1$ or $\mathcal{M}_2 \neq \hat{\mathcal{M}}_2$ tends to zero as $n \to \infty$.

The capacity region $\mathcal{C}(P)$ is defined to be the set of all achievable rate tuples $(R_1(P), R_2(P))$ when the power constraint at the transmitters is P. The DoF region with Shannon feedback is defined as

$$\mathbf{D}^{\mathrm{S}} \stackrel{\triangle}{=} \left\{ (d_1, d_2) \in \mathbb{R}_+^2 \middle| \ \forall \ (w_1, w_2) \in \mathbb{R}_+^2, \right.$$

$$\left. w_1 d_1 + w_2 d_2 \leq \limsup_{P \to \infty} \frac{1}{\log_2 P} \left[\sup_{\left(R_1(P), R_2(P)\right) \in \mathcal{C}(P)} \left\{ w_1 R_1(P) + w_2 R_2(P) \right\} \right] \right\},$$

where $\mathbb{R}_+^2$ denotes the set of pairs of non-negatives real numbers, and $\limsup_{P\to\infty}$ stands for the limit superior [21] as $P \to \infty$. It can be easily proved that the DoF region $\mathbf{D}^{\mathrm{S}}$ is closed [21] and convex [1].

Consider next the MIMO IC under four other settings defined below.

- *Designable Shannon feedback:* both transmitters know the channel matrices $\{H_{ij}(t)\}_{i,j}$ and *modified* channel outputs $\tilde{Y}_1(t)$ and $\tilde{Y}_2(t)$ at time $t+1$ (in general, with some finite delay), where, for each $i \in \{1,2\}$, $\tilde{Y}_i(t) \in \mathbb{C}^{N_i \times 1}$ is a deterministic function of $\overline{Y}_i(t)$ and $\overline{H}(t)$.
- *Limited Shannon feedback of Type 1*: each transmitter knows the other receiver's incoming channels and outputs with some delay (which, without loss of generality we take to be 1 time unit), i.e., the ith transmitter knows the channel matrices $H_{ji}(t)$ and $H_{jj}(t)$ and the received signal $Y_j(t)$ all at time $t+1$, for each $(i,j) \in \{(1,2),(2,1)\}$.
- *Limited Shannon feedback of Type 2*: each transmitter is provided at each time its own receiver's outputs as well as the four channel matrices, all with some delay (which, without loss of generality we take to be 1 time unit); i.e., the ith transmitter knows $Y_i(t)$ and $\{H_{jk}(t)\}_{j,k=1}^2$ at time $t+1$.
- *Output feedback:* both transmitters know the channel outputs $Y_1(t)$ and $Y_2(t)$ at time $t+1$ (or, in general, with a delay of finite number of time slots) but they have no knowledge of channel matrices whatsoever.

The DoF regions of the MIMO IC under these settings can be defined in analogous manner to that with Shannon feedback, and are denoted, respectively, as $\mathbf{D}^{d\mathrm{S}}$, $\mathbf{D}^{l\mathrm{S}1}$, $\mathbf{D}^{l\mathrm{S}2}$, and $\mathbf{D}^{\mathrm{op}}$. Since the designable Shannon feedback setting is stronger than that of Shannon feedback and the limited Shannon feedback and output feedback settings are weaker, we have that

$$\mathbf{D}^{d\mathrm{S}} \supseteq \mathbf{D}^{\mathrm{S}} \supseteq \mathbf{D}^{l\mathrm{S}1}, \qquad \mathbf{D}^{\mathrm{S}} \supseteq \mathbf{D}^{l\mathrm{S}2}, \qquad \mathbf{D}^{\mathrm{S}} \supseteq \mathbf{D}^{\mathrm{op}}.$$

Furthermore, the conditions of delayed CSIT, instantaneous CSIT, and instantaneous CSIT with output feedback are defined as follows:

- *delayed CSIT:* the channel matrices $\{H_{ij}(t)\}_{i,j=1}^2$ are known to the transmitters at time $t+1$;
- *instantaneous CSIT:* the channel matrices $\{H_{ij}(t)\}_{i,j=1}^2$ are known to the transmitters instantaneously (i.e., at time t); and
- *instantaneous CSIT and output feedback:* the channel matrices $\{H_{ij}(t)\}_{i,j=1}^2$ are known to the transmitters at time t, and additionally, they know the channel outputs $Y_1(t)$ and $Y_2(t)$ at time $t+1$.

The DoF regions corresponding to these three assumptions can again be defined analogously, and are denoted, respectively, as $\mathbf{D}^{\mathrm{dCSI}}$, $\mathbf{D}^{\mathrm{iCSI}}$, and $\mathbf{D}^{\mathrm{iCSI\&op}}$. Clearly,

$$\mathbf{D}^{\mathrm{dCSI}} \subseteq \mathbf{D}^{\mathrm{iCSI}}, \qquad \mathbf{D}^{\mathrm{S}} \subseteq \mathbf{D}^{\mathrm{iCSI\&op}}.$$

B. Main Results

The characterization of $\mathbf{D}^{\mathrm{iCSI\&op}}$ below asserts that, in the presence of instantaneous CSIT, output feedback can not improve the DoF region.

Lemma 1: For the MIMO IC with i.i.d. Rayleigh fading, the DoF region with instantaneous CSIT and output feedback is given by

$$\mathbf{D}^{\mathrm{iCSI\&op}} = \Big\{ (d_1, d_2) \Big| 0 \leq d_1 \leq \min(M_1, N_1),\ 0 \leq d_2 \leq \min(M_2, N_2)$$

$$d_1 + d_2 \leq \min\big[M_1 + M_2, N_1 + N_2, \max(M_1, N_2), \max(M_2, N_1) \big] \Big\}.$$

Moreover, $\mathbf{D}^{\mathrm{iCSI}} = \mathbf{D}^{\mathrm{iCSI\&op}}$.

Proof: See Appendix A. ∎

Definition 1: The region $\mathbf{D}_{\text{outer}}^{\text{S}}$ is defined as

$$\mathbf{D}_{\text{outer}}^{\text{S}} \triangleq \left\{ (d_1, d_2) \middle| \; L_{0i} \triangleq 0 \leq d_i \leq \min(M_i, N_i), \; i = 1, 2; \right.$$

$$L_1 \triangleq \frac{d_1}{\min(N_1 + N_2, M_1)} + \frac{d_2}{\min(N_2, M_1)} \leq \frac{\min(N_2, M_1 + M_2)}{\min(N_2, M_1)};$$

$$L_2 \triangleq \frac{d_1}{\min(N_1, M_2)} + \frac{d_2}{\min(N_1 + N_2, M_2)} \leq \frac{\min(N_1, M_1 + M_2)}{\min(N_1, M_2)};$$

$$\left. L_3 \triangleq d_1 + d_2 \leq \min\left[M_1 + M_2, N_1 + N_2, \max(M_1, N_2), \max(M_2, N_1)\right] \right\}.$$

Note that the first two bounds on d_1 and d_2 have been denoted as L_{01} and L_{02} respectively; while the last three bounds on the weighted sums of d_1 and d_2 are denoted respectively by L_1, L_2, and L_3.

Theorem 1 (Outer-Bound with Shannon feedback): For the MIMO IC with i.i.d. Rayleigh fading, the DoF region with Shannon feedback is outer-bounded by the region $\mathbf{D}_{\text{outer}}^{\text{S}}$, i.e.,

$$\mathbf{D}^{\text{S}} \subseteq \mathbf{D}_{\text{outer}}^{\text{S}}.$$

Proof: See Section III. ∎

Definition 2: For a given $i \in \{1, 2\}$, Condition i is said to hold whenever the inequality

$$M_i > N_1 + N_2 - M_j > N_i > N_j > M_j > N_j \frac{N_j - M_j}{N_i - M_j}$$

holds for $j \in \{1, 2\}$ with $j \neq i$.

Clearly, the two conditions are symmetric counterparts of each other. Moreover, the two conditions can not be true simultaneously. Condition i can not hold if $N_j \geq N_i$.

Theorem 2 (The DoF Region with Shannon feedback): For the MIMO IC with i.i.d. Rayleigh fading, the DoF region with Shannon feedback is equal to the region $\mathbf{D}_{\text{outer}}^{\text{S}}$, i.e.,

$$\mathbf{D}^{\text{S}} = \mathbf{D}_{\text{outer}}^{\text{S}}.$$

Proof: It is sufficient to prove that the region $\mathbf{D}_{\text{outer}}^{\text{S}}$ is achievable when there is Shannon feedback. We assume without loss of generality that $N_1 \geq N_2$ (note, under this assumption, that Condition 2 can not hold).

Suppose Condition 1 does not hold. Then from [20, Theorem 2], we observe that

$$\mathbf{D}_{\text{outer}}^{\text{S}} = \mathbf{D}^{\text{dCSI}},$$

so that the theorem follows by noting that $\mathbf{D}^{\text{dCSI}} \subseteq \mathbf{D}^{\text{S}} \subseteq \mathbf{D}_{\text{outer}}^{\text{S}}$.

Thus, it is only required to show that the region $\mathbf{D}_{\text{outer}}^{\text{S}}$ is achievable when Condition 1 holds. The detailed proof is given in Section IV. ∎

The basic idea behind the interference alignment (IA) based achievability scheme developed in Section IV to prove the above theorem is illustrated via an example in Section II-C which shows that $\mathbf{D}^{\text{S}} \neq \mathbf{D}^{\text{dCSI}}$ and provides insight as to why the DoF regions are not always identical. Further, Section II-D compares the techniques used to achieve IA under Shannon feedback and under delayed CSIT.

Remark 1 (Comparison of $\mathbf{D}^{\text{S}}$ and $\mathbf{D}^{\text{dCSI}}$): Using Theorem 2 above and [20, Theorem 2], we observe that $\mathbf{D}^{\text{S}} \neq \mathbf{D}^{\text{dCSI}}$ if and only if Conditions 1 or 2 hold. In other words, in the presence of delayed CSIT, output feedback helps in improving the DoF region only when Conditions 1 or 2 hold.

Remark 2: Using Lemma 1, we observe that output feedback can not enhance the DoF region when there is instantaneous CSIT. In contrast, output feedback improves the DoF region when there is just delayed CSIT.

The next two corollaries extend the above results to MIMO ICs with limited Shannon feedback of Type I and Type II, just output feedback and with designable Shannon feedback.

Corollary 1: For the MIMO IC with i.i.d. Rayleigh fading, we have

$$\mathbf{D}^{l\mathrm{S}1} = \mathbf{D}^{l\mathrm{S}2} = \mathbf{D}^{d\mathrm{S}} = \mathbf{D}^{\mathrm{S}}.$$

Proof: See Appendix B-A. ∎

Corollary 2: For the MIMO IC with i.i.d. Rayleigh fading, we have

$$\mathbf{D}^{\mathrm{op}} = \mathbf{D}^{\mathrm{S}},$$

if neither of the following two inequalities hold: $\min(M_1, N_1) > N_2 > M_2$ and $\min(M_2, N_2) > M_1 > N_1$.

Proof: See Appendix B-B ∎

Thus, the above corollary yields the DoF region with output feedback for a large class of MIMO ICs. When one of the above two conditions holds the DoF region with output feedback is not known.

Following the submission of a conference version of this work, and simultaneously with its publication in [22], Tandon et. al. reported the DoF region for limited Shannon feedback of Type II in [23].

C. Retrospective Interference Alignment with Shannon Feedback

With Theorem 2 in hand, we know that $\mathbf{D}^{\mathrm{S}} \supset \mathbf{D}^{\mathrm{dCSI}}$ in general. However, the proof of this theorem is involved, and therefore, we provide an example in which $\mathbf{D}^{\mathrm{S}} \neq \mathbf{D}^{\mathrm{dCSI}}$ by demonstrating that a point outside $\mathbf{D}^{\mathrm{dCSI}}$ can be achieved with Shannon feedback. The proof that this scheme works is based on a series of simple propositions.

In particular, we consider the $(6, 2, 4, 3)$ MIMO IC shown in Fig. 3. For this IC, the DoF regions with just delayed CSIT and with Shannon feedback are plotted in Fig. 2, from which we observe that the former is strictly smaller than the latter. Moreover, it is clear from Fig. 2 that when $d_2 = 2$, $d_1 \leq \frac{5}{3}$ with delayed CSIT. Here, we prove the achievability the DoF pair $(\frac{12}{7}, 2)$ with Shannon feedback, which establishes that $\mathbf{D}^{\mathrm{S}} \neq \mathbf{D}^{\mathrm{dCSI}}$ since $\frac{12}{7} > \frac{5}{3}$.

Toward this end, we show that by coding over 7 times slots, 12 and 14 DoF can be achieved for the first and the second transmit-receive pairs, respectively. In our scheme, T2 transmits 2 data symbols (DSs) intended for R2 over each time slot and thus a total of 14 DSs are sent; whereas T1 transmits 6 DSs intended for R1 at $t = 1$ and $t = 4$. Further, at $t = 7$, we show that desired DSs can be successfully decoded by both receivers.

Consider first a transformation which simplifies the description of our scheme. At time t, the ith receiver can compute a unitary matrix $U_{i2}(t)$ such that it is deterministic function of $H_{i2}(t)$ and the bottom $(N_i - 2)$ rows of the transformed matrix $U_{i2}(t)H_{i2}(t)$ consist only of zeros. Using it, the ith receiver evaluates the transformed output $U_{i2}(t)Y_i(t)$. Henceforth, the transformed quantities $U_{i2}(t)H_{i2}(t)$ and $U_{i2}(t)Y_i(t)$ are denoted simply as $H_{i2}(t)$ and
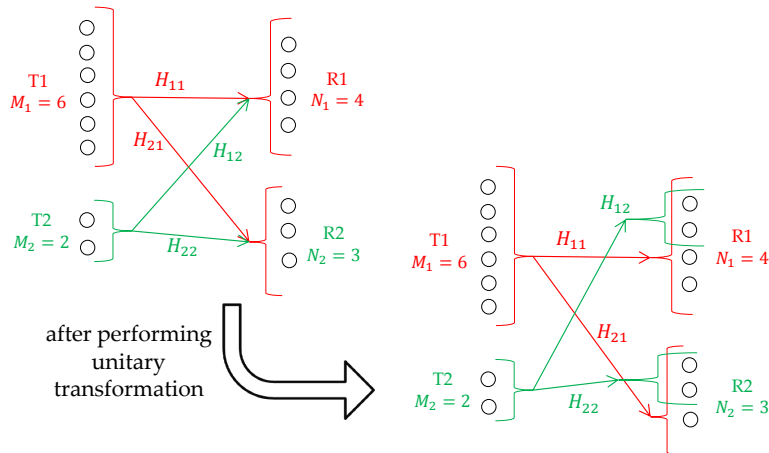
It is assumed, without loss of generality, that only the first two rows of H_{12} and H_{22} contain non-zero entries.

Fig. 3. The $(6, 2, 4, 3)$ MIMO IC Considered in Section II-C
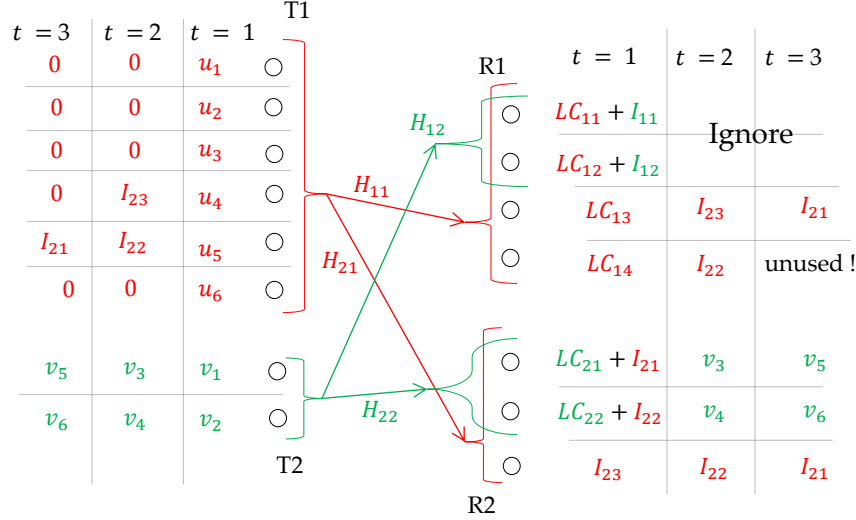
Fig. 4. IA scheme for Achieving $(\frac{12}{7}, 2)$ with Shannon Feedback over the $(6, 2, 4, 3)$ MIMO IC: $t = 1$ to $t = 3$

$Y_i(t)$, respectively. Evidently, the transmit signal $X_2(t)$ affects only the first two entries of (the transformed) $Y_i(t)$. Hence, throughout this subsection, we assume without loss of generality that the bottom $(N_i - 2)$ rows of $H_{i2}(t)$ consist only of zeros for all t, and thus, the signal $X_2(t)$ can affect only the first two antennas of R1 and R2 (see also Fig. 3).

Consider the operation of the scheme at $t = 1$; see also Fig. 4. At this time, T1 transmits i.i.d. complex Gaussian DSs $\{u_i\}_{i=1}^{6}$ intended for R1, while T2 sends i.i.d. complex Gaussian DSs v_1 and v_2 for R2. Thus, the transmit signals are formed (for a vector V_i, V_{ij} denotes its jth entry) as follows: $X_{1i}(1) = u_i \ \forall \ i \in [1 : 6]$ and $X_{2j}(1) = v_j$, $j = 1, 2$. The received signals at R1 and R2 can be written as follows (note that since the additive noises do not alter a DoF result, they are ignored) with desired and interfering linear combinations of data symbols defined using the symbols LC and I, respectively, so that

$$
\begin{aligned}
Y_{1i}(1) &= H_{1i1}(1)X_1(1) + H_{1i2}(1)X_2(1), & i \in [1 : 4] \\
&= \underbrace{H_{1i1}(1)\begin{bmatrix} u_1^* & u_2^* & \cdots & u_6^* \end{bmatrix}^*}_{\triangleq \text{LC}_{1i}} + \underbrace{H_{1i2}(1)\begin{bmatrix} v_1^* & v_2^* \end{bmatrix}^*}_{\triangleq \text{I}_{1i}} \\
Y_{2i}(1) &= H_{2j2}(1)X_2(1) + H_{2j1}(1)X_1(1) & j \in [1 : 2] \\
&= \underbrace{H_{2j2}(1)\begin{bmatrix} v_1^* & v_2^* \end{bmatrix}^*}_{\triangleq \text{LC}_{2j}} + \underbrace{H_{2j1}(1)\begin{bmatrix} u_1^* & u_2^* & \cdots & u_6^* \end{bmatrix}^*}_{\triangleq \text{I}_{2j}},
\end{aligned}
$$

and moreover, since we have assumed without loss of generality that the bottom $(N_i - 2)$ rows of $H_{i2}(t)$ consist only of zeros $\forall \ t$ and i, we have $\text{I}_{13} = \text{I}_{14} = \text{LC}_{23} = 0$ (see also Fig. 4). Thus,

$$
\begin{aligned}
Y_{1i}(1) &= \text{LC}_{1i} + \text{I}_{1i} & i \in [1 : 2], \\
Y_{1i}(1) &= \text{LC}_{1i} & i \in [3 : 4], \\
Y_{2j}(1) &= \text{LC}_{2j} + \text{I}_{2j} & j \in [1 : 2], \quad \text{and} \\
Y_{23}(1) &= \text{I}_{23}.
\end{aligned}
$$

At time $t = 1$, both receivers encounter interference, and therefore, can not decode their desired data symbols. Moreover, the interference at a given antenna of a receiver is the linear combination of the DSs sent by its unpaired

transmitter. Thus, each transmitter can compute the past interference encountered by its unpaired receiver using just delayed CSI, as stated in the following proposition.

Proposition 1: T1, at time $t = 2$, can compute $I_{2j} \; \forall \; j \in [1 : 3]$ using just delayed CSI.

Proof: T1 knows DSs u_i's; and time $t = 2$, and it also knows $H_{2j1}(1)$ because of delayed CSIT knowledge and hence $I_{2j} = H_{2j1}(1) \begin{bmatrix} u_1^* & u_2^* & \cdots & u_6^* \end{bmatrix}^*$. ∎

Hence, at $t = 2$ and $t = 3$, T1 transmits $I_{2j} \; \forall \; j \in [1 : 3]$ as shown in Fig. 4, while T2 continues to transmit 2 new DSs intended for R2. In particular, the transmit signals are given as

$$
\begin{aligned}
X_{1i}(2) &= X_{1i}(3) = 0 \; \forall \; i \in [1 : 3], \\
X_{14}(2) &= I_{23}, \qquad X_{15}(2) = I_{22}, \qquad X_{16}(2) = 0 \\
X_{14}(3) &= I_{23}, \qquad X_{15}(3) = X_{16}(3) = 0 \\
X_{21}(2) &= v_3, \qquad X_{22}(2) = v_4 \\
X_{21}(3) &= v_5, \qquad X_{22}(3) = v_6,
\end{aligned}
$$

where $v_3, \cdots, v_6$ are i.i.d. complex Gaussian DSs intended for R2. Consider now the decoding operation at the receivers, starting with R2. The following proposition states that R2 can decode the desired DSs $v_1, \cdots, v_6$ at time $t = 3$.

Proposition 2: At time $t = 2$, R2 can decode v_3, v_4, and I_{22}. At time $t = 3$, R2 can decode v_5, v_6, and I_{21}. After determining I_{22} and I_{21}, R2 can evaluate v_1 and v_2, and thus, at $t = 3$, decoding is successful at R2.

Proof: At time $t = 1$, R2 knows I_{23}. Hence, it can subtract the contribution due to I_{23} from $Y_2(2)$. Thus, equivalently, for R2, only 3 transmit antennas sending a non-zero signal at this time. Therefore, R2, via simple channel inversion, can determine v_3, v_4, and I_{22}. That at time $t = 3$, R2 can decode v_5, v_6, and I_{21} follows similarly. After knowing the values of I_{21} and I_{22}, R2 can evaluate $Y_{21}(1) - I_{21} = \text{LC}_{21}$, and similarly, LC_{22}. In other words, at $t = 3$, it can obtain 2 linear combinations (LCs) of v_1 and v_2. Thus, it can decode v_1 and v_2. ∎

Consider now the case of R1. At $t = 3$, as per the next proposition, R1 knows I_{21}, I_{22}, and I_{23}. Since these are linear combinations of $u_1, \cdots, u_6$, they are useful for R1.

Proposition 3: R1 can determine the values of I_{22} and I_{23} at time $t = 2$, and that of I_{21} at time $t = 3$.

Proof: R1 can simply ignore the first two receive antennas, which experience interference due to the signal of T2. Then, using the last two antennas, it can compute the required symbols using channel inversion. ∎

Thus, at $t = 3$, R1 gets 5 linear combinations, namely, LC_{23}, LC_{24}, I_{21}, I_{22}, and I_{23}, of 6 desired DSs. Thus, it needs one more useful linear combination for successful decoding. Consider the next proposition.

Proposition 4: Given that R1 knows the values of LC_{23}, LC_{24}, I_{21}, I_{22}, and I_{23}, it can decode the six symbols u_1, u_2, $\cdots$, u_6, provided it knows I_{12}.

Proof: If R1 knows LC_{23}, LC_{24}, I_{21}, I_{22}, and I_{23}, then it is sufficient for it to know one more linear combination of u_1, u_2, $\cdots$, u_6, which it can compute using I_{12} as follows: $Y_{12}(1) - I_{12} = \text{LC}_{12}$ is a linear combination of R1's desired symbols. ∎

Hence, it is sufficient to communicate the value of I_{12} to R1. Consider the following proposition.

Proposition 5: T1 can compute I_{12} at $t = 2$ using Shannon feedback, but not using just delayed CSI. Moreover, R2 knows I_{12} at time $t = 3$.
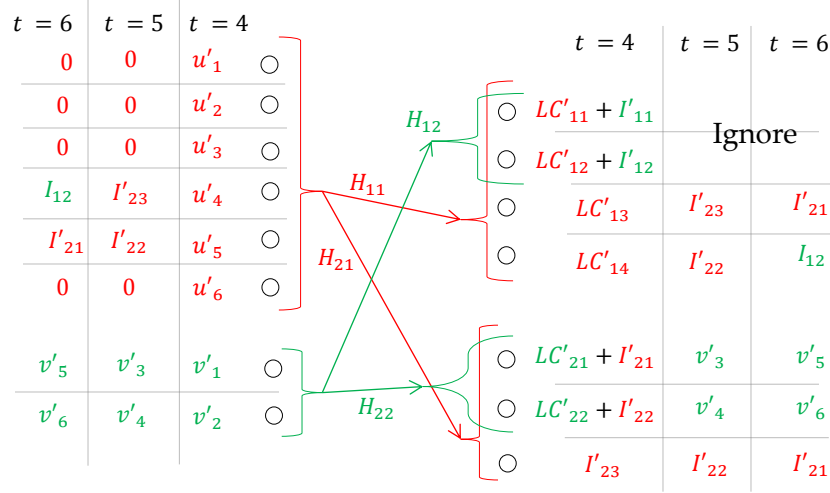
Proof: Because of Shannon feedback, T1 knows $Y_2(1)$, $H_{21}(1)$, and $H_{22}(1)$ at time $t = 2$ by virtue of Shannon feedback. Since it knows $X_1(1)$ by default, it can compute

$$
\left(H_{22}(1) \right)^{\dagger} \left\{ Y_2(1) - H_{21}(1) X_1(1) \right\} = \left(H_{22}(1) \right)^{\dagger} H_{22}(1) X_2(1) = X_2(1)
$$

at time $t = 2$, where $\left(H_{22}(1) \right)^{\dagger}$ is the pseudo-inverse of $H_{22}(1)$ [24]. Subsequently, it can evaluate I_{12}, which is a linear combination of v_1 and v_2. R2, after decoding v_1 and v_2 at time $t = 3$, can compute I_{12} since it knows all channel matrices. ∎

Thus, in light of this proposition, T1 can convey I_{12} to R1 at time $t = 6$ without interfering with decoding at R2.

Consider the next block of 3 time slots, i.e., for $t = 4, 5, 6$. The scheme here is identical to that for the first three time slots; see Fig. 5. At time $t = 4$, T1 transmits DSs u_1', u_2', $\cdots$, u_6' intended for R1. T2, on the other hand, transmits DSs v_1', v_2', $\cdots$, v_6' for R2. Here, the superscript prime is used to indicate the quantities that are specific

$t=6$	$t=5$	$t=4$			$t=4$	$t=5$	$t=6$
0	0	u'_1			$LC'_{11}+I'_{11}$	Ignore	
0	0	u'_2			$LC'_{12}+I'_{12}$		
0	0	u'_3			LC'_{13}	I'_{23}	I'_{21}
I_{12}	I'_{23}	u'_4			LC'_{14}	I'_{22}	I_{12}
I'_{21}	I'_{22}	u'_5					
0	0	u'_6					
v'_5	v'_3	v'_1			$LC'_{21}+I'_{21}$	v'_3	v'_5
v'_6	v'_4	v'_2			$LC'_{22}+I'_{22}$	v'_4	v'_6
					I'_{23}	I'_{22}	I'_{21}

(H_{12}, H_{11}, H_{21}, H_{22})

- R1 can decode its desired symbols if it knows I'_{12}.
- R2, after decoding v'_1 and v'_2, can compute I'_{12}.
- T1 knows v'_1 and v'_2 at $t=5$, and hence also I'_{12}.

Fig. 5.   IA scheme for Achieving $\left(\frac{12}{7},2\right)$ with Shannon Feedback over the $(6,2,4,3)$ MIMO IC: $t=4$ to $t=6$

$t=7$			$t=7$
0			Ignore
0			
0			
0			
0			I'_{12}
I'_{12}			
v''_1			v''_1
v''_2			v''_2

(H_{12}, H_{11}, H_{21}, H_{22})
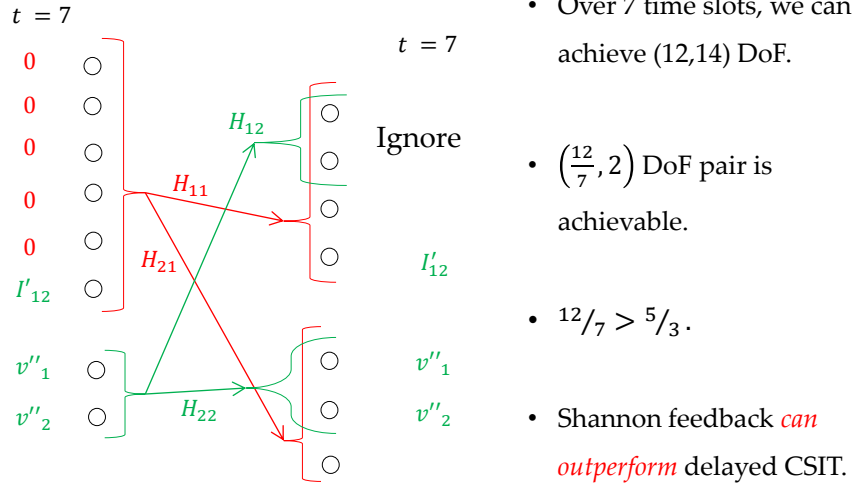
- Over 7 time slots, we can achieve (12,14) DoF.
- $\left(\frac{12}{7},2\right)$ DoF pair is achievable.
- $^{12}/_7 > ^{5}/_3$.
- Shannon feedback *can outperform* delayed CSIT.

Fig. 6.   IA scheme for Achieving $\left(\frac{12}{7},2\right)$ with Shannon Feedback over the $(6,2,4,3)$ MIMO IC: $t=7$

to this block of three time slots. The only change in the transmission scheme is that at time $t=6$, T1 transmits I_{12}. Consider the following propositions which will be used to describe decoding at the receivers.

Proposition 6: Consider receiver R2: (a) at time $t=5$, R2 can decode v'_3, v'_4, and I'_{22} (b) at time $t=6$, R2 can decode v'_5, v'_6, and I'_{21} and (c) after determining I'_{22} and I'_{23}, R2 can evaluate v'_1 and v'_2, and thus, at $t=6$, R2 can perform successful decoding.

Proof: The proofs of Parts (a) and (c) similar to those of Parts (i) and (iii) of Proposition 2. Further, Part (b) follows from the proof of Proposition 2(ii) by noting that I_{12} is known to R1 at time $t=3$. ∎

Proposition 7: Consider receiver R1: (a) R1 knows I_{12} at time $t=6$, from which it can decode $u_1, u_2, \cdots, u_6$, (b) R1 can compute I'_{22} and I'_{23} at time $t=5$, and I'_{21} at time $t=6$ (c) if R1 is conveyed the value of I'_{12} at time $t=7$, it can decode $u'_1, u'_2, \cdots, u'_6$ and (d) T1 knows I'_{12} at time $t=5$.

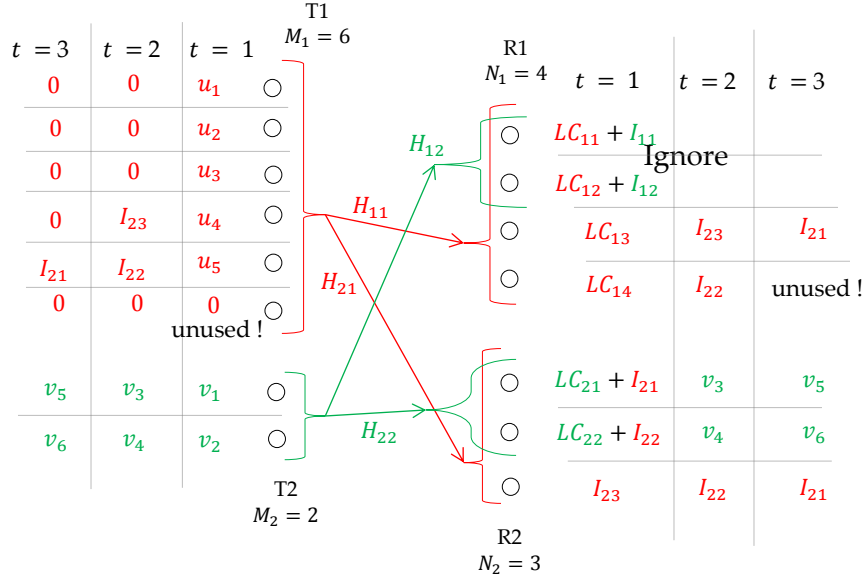*Proof:* The four parts follow respectively from Propositions 4, 3, 4, and 5. ∎

Fig. 7. IA scheme for Achieving $\left(\frac{5}{3}, 2\right)$ with Delayed CSIT over the $(6, 2, 4, 3)$ MIMO IC

Thus, as per the two propositions, at $t = 6$, R2 can decode all symbols sent to it until that time, whereas R1 can do the same if it is delivered the value of I'_{12} at time $t = 7$. Next, consider $t = 7$. As shown in Fig. 6, T1 transmits just I'_{12}, while T2 sends two new data symbols v''_1 and v''_2. It is easy to show that R1 can decode I'_{12}, whereas R2 can decode v''_1 and v''_2. Hence, as desired, we can achieve a DoF pair $(12, 14)$ over 7 symbol times.

It is instructive to compare the above Shannon feedback scheme with the delayed-CSIT coding scheme of [20] that can only achieve the pair $\left(\frac{5}{3}, 2\right)$ over the $(6, 2, 4, 3)$ MIMO IC as illustrated in Fig. 7 in terms of the notation introduced earlier in this sub-section. In this latter case, by coding over 3 time slots, we achieve 5 and 6 DoF for the two transmit-receive pairs, respectively. Note that T1 can not determine I_{12} (and I'_{12}) with just delayed CSIT, and thus only 5 linear combinations can be delivered to R1 over a span of 3 time slots. Hence, T1 transmits only 5 DSs for R1 at time $t = 1$, which R1 can decode at $t = 3$. Except for this difference (compare Figs. 4 and 7), this coding scheme is identical to the Shannon feedback coding scheme.

Note that under this delayed-CSIT scheme, the sixth antenna of R1 is never used. Moreover, one of the last two antennas of R1, say, the fourth, remains unused, although it never experiences interference. In other words, in the delayed-CSIT scheme, some of the resources are not utilized. Shannon feedback on the other hand permits the exploitation of these resources – the sixth antenna of T1 is used with Shannon feedback and both interference-free antennas of R1 are used under the Shannon-feedback scheme at $t = 4, 5, 6$ – thereby outperforming delayed CSIT feedback.

D. Comparison of IA with Shannon Feedback and IA with Delayed CSIT

In the Shannon-feedback coding scheme of the previous sub-section (and more generally of Section IV), one may observe that the following two types of techniques are used to achieve IA:

1) Since the interference at a given receiver is a linear combinations of the DSs sent by its unpaired transmitter, each transmitter, using delayed CSIT, can evaluate and then transmit the interference seen in the past by its *unpaired* receiver to convey new useful linear combinations to its paired receiver without creating any new additional interference to the unpaired receiver.

2) Equipped with the knowledge of past channel outputs and past channel matrices, each transmitter can compute and transmit the interference encountered in the past by its *paired* receiver to provide an opportunity to its paired receiver to cancel the past interference while conveying useful information to its unpaired receiver.

Note that the DoF-region-optimal IA-based achievability schemes developed in [20] for the MIMO IC with *just* delayed CSIT make use of the first technique but not the second one, because the latter is feasible only in the presence of output feedback. Output feedback with delayed CSIT on the other hand enables each transmitter to
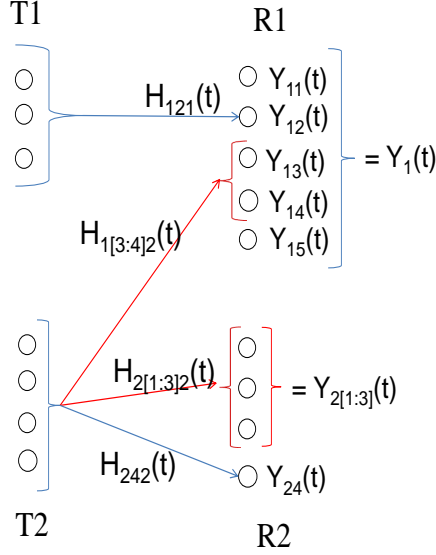
Fig. 8. Illustration of notation.

compute the past transmit signal of the other transmitter which introduces partial transmitter cooperation which is infeasible when there is just delayed CSIT; remarkably, this transmitter cooperation reveals all available signaling dimensions and achieves the DoF gains promised by Shannon feedback.

Note that while transmitter cooperation is induced by output feedback regardless of whether there is instantaneous CSIT or delayed CSIT, it is only in the case of delayed CSIT that such cooperation provides a DoF-region improvement. With instantaneous CSIT, the transmit signals can be suitably beamformed to cause minimal interference at the receivers. With delayed CSIT however, transmit beamforming can not be employed and hence the receivers experience a relatively high amount of interference so that output feedback is more effective.

III. PROOF OF THEOREM 1

If $(d_1, d_2) \in \mathbf{D}^{\mathrm{S}}$, then $(d_1, d_2) \in \mathbf{D}^{\mathrm{iCSI\&op}}$. Therefore, by Lemma 1, bounds L_{01}, L_{02}, and L_3 must hold at any $(d_1, d_2) \in \mathbf{D}^{\mathrm{S}}$. Now, note that L_1 and L_2 are symmetric counterparts of each other (i.e., any one of them can be obtained from the other by changing the user ordering). Hence, it is sufficient to prove that L_1 holds, which is the goal of the remainder of this section.

Before we prove that L_1 is an outer-bound, we introduce some further notation used henceforth in the paper.

Notation: The set of four channel matrices at time t is denoted by $H(t)$, i.e., $H(t) = \{H_{ij}(t)\}$ where $i, j \in \{1, 2\}$. For integers n_1 and n_2, if $n_1 \le n_2$, $[n_1 : n_2] = \{n_1, n_1 + 1, \cdots, n_2\}$; whereas if $n_1 > n_2$, then $[n_1 : n_2]$ denotes the empty set. For a random variable $X(t)$, $X([n_1 : n_2]) = \{X(t)\}_{t=n_1}^{n_2}$ if $n_1 \le n_2$, whereas $X([n_1 : n_2])$ denotes an empty set if $n_1 > n_2$. Further, for $n \ge 1$, $\overline{X}(n) = X([1 : n])$. For the received signal $Y_i(t)$ and the channel matrix $H_{ik}(t)$, the jth entry and the jth row are denoted respectively by $Y_{ij}(t)$ and $H_{ijk}(t)$. Further, whenever $n_1 \le n_2$ and $n_3 \le n_4$, $Y_{i[n_1:n_2]}(t) = \{Y_{ij}(t)\}_{j=n_1}^{n_2}$, $Y_{i[n_1:n_2]}([n_3 : n_4]) = \{\{Y_{ij}(t)\}_{j=n_1}^{n_2}\}_{t=n_3}^{n_4}$, $H_{i[n_1:n_2]j}(t)$ is the channel matrix from j^{th} transmitter to channel outputs $Y_{i[n_1:n_2]}(t)$ (see Fig. 8); however, if $n_1 > n_2$ and/or $n_3 > n_4$, then $Y_{i[n_1:n_2]}(t)$ and $Y_{i[n_1:n_2]}([n_3 : n_4])$ denote empty sets. Moreover, for $n \ge 1$, $\overline{Y}_{i[n_1:n_2]}(n) = Y_{i[n_1:n_2]}([1 : n])$. Finally, $o(\log_2 P)$ denotes any real-valued function $x(P)$ of P such that $\lim_{P \to \infty} \frac{x(P)}{\log_2 P} = 0$.

We will show that the bound L_1 must hold at any $(d_1, d_2) \in \mathbf{D}^{\mathrm{S}}$. We first apply Fano's inequality to upper-bound

the rates achievable for the two users starting below with R_2.

$$nR_2 \leq I\Big(\mathcal{M}_2; \overline{Y}_2(n) \Big| \overline{H}(n)\Big) + n\epsilon_n \tag{3}$$

$$= h\Big(\overline{Y}_2(n) \Big| \overline{H}(n)\Big) - h\Big(\overline{Y}_2(n) \Big| \mathcal{M}_2, \overline{H}(n)\Big) + n\epsilon_n \tag{4}$$

$$= h\Big(\overline{Y}_2(n) \Big| \overline{H}(n)\Big) - \sum_{t=1}^{n} h\Big(Y_2(t) \Big| \overline{Y}_2(t-1), \mathcal{M}_2, \overline{H}(n)\Big) + n\epsilon_n \tag{5}$$

$$\leq h\Big(\overline{Y}_2(n) \Big| \overline{H}(n)\Big) - \sum_{t=1}^{n} h\Big(Y_2(t) \Big| \overline{Y}_2(t-1), \mathcal{M}_2, \overline{X}_2(t), \overline{H}(n)\Big) + n\epsilon_n, \tag{6}$$

$$= h\Big(\overline{Y}_2(n) \Big| \overline{H}(n)\Big) - \sum_{t=1}^{n} h\Big(Y_2(t) \Big| \overline{Y}_2(t-1), \mathcal{M}_2, \overline{X}_2(t), \overline{H}(t)\Big) + n\epsilon_n, \tag{7}$$

where $\epsilon_n \to 0$ as $n \to \infty$; the inequality (6) holds since conditioning reduces entropy [1]; and the equality in (7) follows on noting that random variables $\{Y_2(t), \overline{Y}_2(t-1), \mathcal{M}_2, \overline{X}_2(t)\}$ are independent of $H([t+1:n])$.

We next use Fano's inequality at R1 assuming that it knows the received signal $Y_2(t)$ instantaneously and also the message $\mathcal{M}_2$ to obtain the following:

$$nR_1 \leq I\Big(\mathcal{M}_1; \overline{Y}_2(n), \overline{Y}_1(n), \mathcal{M}_2 \Big| \overline{H}(n)\Big) + n\epsilon_n$$

$$= I\Big(\mathcal{M}_1; \overline{Y}_1(n), \overline{Y}_2(n) \Big| \mathcal{M}_2, \overline{H}(n)\Big) + n\epsilon_n \tag{8}$$

$$= \sum_{t=1}^{n} h\Big(Y_1(t), Y_2(t) \Big| \overline{Y}_1(t-1), \overline{Y}_2(t-1), \mathcal{M}_2, \overline{H}(n)\Big)$$
$$- \sum_{t=1}^{n} h\Big(Y_1(t), Y_2(t) \Big| \overline{Y}_1(t-1), \overline{Y}_2(t-1), \mathcal{M}_2, \mathcal{M}_1, \overline{H}(n)\Big) + n\epsilon_n \tag{9}$$

$$= \sum_{t=1}^{n} h\Big(Y_1(t), Y_2(t) \Big| \overline{Y}_1(t-1), \overline{Y}_2(t-1), \mathcal{M}_2, \overline{X}_2(t), \overline{H}(n)\Big)$$
$$- \sum_{t=1}^{n} h\Big(Y_1(t), Y_2(t) \Big| \overline{Y}_1(t-1), \overline{Y}_2(t-1), \mathcal{M}_2, \mathcal{M}_1, \overline{X}_1(t), \overline{X}_2(t), \overline{H}(n)\Big) + n\epsilon_n \tag{10}$$

$$= \sum_{t=1}^{n} h\Big(Y_1(t), Y_2(t) \Big| \overline{Y}_1(t-1), \overline{Y}_2(t-1), \mathcal{M}_2, \overline{X}_2(t), \overline{H}(t)\Big)$$
$$- \sum_{t=1}^{n} h\Big(Y_1(t), Y_2(t) \Big| \overline{Y}_1(t-1), \overline{Y}_2(t-1), \mathcal{M}_2, \mathcal{M}_1, \overline{X}_1(t), \overline{X}_2(t), \overline{H}(t)\Big) + n\epsilon_n \tag{11}$$

$$= \sum_{t=1}^{n} h\Big(Y_1(t), Y_2(t) \Big| \overline{Y}_1(t-1), \overline{Y}_2(t-1), \mathcal{M}_2, \overline{X}_2(t), \overline{H}(t)\Big)$$
$$- \sum_{t=1}^{n} h\Big(W_1(t), W_2(t)\Big) + n\epsilon_n \tag{12}$$

$$= \sum_{t=1}^{n} \Big\{ h\Big(Y_1(t), Y_2(t) \Big| \overline{Y}_1(t-1), \overline{Y}_2(t-1), \mathcal{M}_2, \overline{X}_2(t)\overline{H}(t)\Big) + o(\log_2 P) + \epsilon_n \Big\}, \tag{13}$$

where various steps follow because of the following reasons: the equality in (8) holds due to the independence of the two messages; equality (9) holds because of the definition of the mutual information and the chain rule for the differential entropy; equality (10) follows by noting that the transmit signal $X_i(t)$ is a deterministic function of $\mathcal{M}_i$, $\overline{Y}_1(t-1)$, $\overline{Y}_2(t-1)$, and $\overline{H}(t)$; (11) holds since all the involved random variables are independent of $H([t+1:n])$; (12) holds because translation does not change differential entropy, and $W_1(t)$ and $W_2(t)$ are independent of $\overline{Y}_1(t-1)$, $\overline{Y}_2(t-1)$, $\mathcal{M}_2$, $\mathcal{M}_1$, $\overline{X}_1(t)$, $\overline{X}_2(t)$, and $\overline{H}(t)$; the final equality holds since the noises are i.i.d. across time and their statistics are independent of P.

Lemma 2: Let $m_1 \triangleq \min(M_1, N_1 + N_2)$ and $m_2 \triangleq \min(M_1, N_2)$. Then, for each $t \in [1 : n]$, we have

$$\frac{1}{m_2} h\left(Y_2(t) \Big| \overline{Y}_2(t-1), \mathcal{M}_2, \overline{X}_2(t), \overline{H}(t)\right)$$

$$\geq \frac{1}{m_1} h\left(Y_1(t), Y_2(t) \Big| \overline{Y}_1(t-1), \overline{Y}_2(t-1), \mathcal{M}_2, \overline{X}_2(t), \overline{H}(t)\right) + o(\log_2 P)$$

where the term $o(\log_2 P)$ is constant with n.

Proof: See Section III-A. ∎

Combining the bounds in (7), (13), and the one in Lemma 2, we get

$$\frac{1}{m_2} R_2 \leq \frac{1}{m_2 \cdot n} h\left(\overline{Y}_2(n) \Big| \overline{H}(n)\right) + \frac{\epsilon_n}{m_2} - \left\{\frac{1}{m_1} R_1 - o(\log_2 P) - \epsilon_n\right\}$$

$$\Rightarrow \frac{R_2}{m_2} + \frac{R_1}{m_1} \leq \frac{\min(N_2, M_1 + M_2)}{m_2} \cdot \log_2 P + \epsilon_n \left(\frac{1}{m_2} + 1\right) + o(\log_2 P),$$

where the last inequality holds since the DoF of the point-to-point MIMO channel are equal to the minimum of the number of transmit and receive antennas. Since $\epsilon_n \to 0$ as $n \to \infty$, we now have

$$\frac{R_2}{m_2} + \frac{R_1}{m_1} \leq \frac{\min(N_2, M_1 + M_2)}{m_2} \cdot \log_2 P + o(\log_2 P)$$

$$\Rightarrow \frac{d_2}{m_2} + \frac{d_1}{m_1} \leq \limsup_{P \to \infty} \frac{R_2}{m_2} + \frac{R_1}{m_1} \leq \frac{\min(N_2, M_1 + M_2)}{m_2}$$

as desired.

A. Proof of Lemma 2

In the following two lemmas, it is shown that although the received signals $Y_1(t)$ and $Y_2(t)$ are N_1 and N_2 dimensional, respectively, only the first $m_1 - m_2$ and m_2 entries of them are relevant as far as the current DoF analysis is concerned.

Lemma 3: If $m_2 = \min(M_1, N_2)$, we have the following:

$$h\left(Y_2(t) \Big| \overline{Y}_2(t-1), \mathcal{M}_2, \overline{X}_2(t), \overline{H}(t)\right) \geq h\left(Y_{2[1:m_2]}(t) \Big| \overline{Y}_2(t-1), \mathcal{M}_2, \overline{X}_2(t), \overline{H}(t)\right) + o(\log_2 P),$$

where the term $o(\log_2 P)$ is constant with n.

Proof: Follows from the techniques in [20, Proof of Lemma 2]. ∎

Lemma 4: If $m_1 = \min(M_1, N_1 + N_2)$, then

$$h\left(Y_1(t), Y_2(t) \Big| \overline{Y}_1(t-1), \overline{Y}_2(t-1), \mathcal{M}_2, \overline{X}_2(t), \overline{H}(t)\right)$$

$$\leq h\left(Y_{1[1:m_1-m_2]}(t), Y_{2[1:m_2]}(t) \Big| \overline{Y}_1(t-1), \overline{Y}_2(t-1), \mathcal{M}_2, \overline{X}_2(t), \overline{H}(t)\right) + o(\log_2 P).$$

where the term $o(\log_2 P)$ is constant with n.

Proof: Follows from the techniques in [20, Proof of Lemma 3]. ∎

If $m_1 - m_2 = 0$, Lemma 2 holds trivially. Hence, in the following, we may consider without loss of generality that $m_1 > m_2$.

We now prove the following lemma which is critical in the proof of Lemma 2.

Lemma 5: Let $Q(t) \triangleq \{\mathcal{M}_2, \overline{H}(t), \overline{Y}_2(t-1), \overline{X}_2(t)\}$. For an $i \in [1 : m_2 - 1]$ and a $k \in [1 : m_1 - m_2]$, if $j = i + 1$ and $l = k + 1$, we have the following equalities:

$$h\left(Y_{2i}(t) \Big| Q(t), Y_{2[1:i-1]}(t)\right) = h\left(Y_{2j}(t) \Big| Q(t), Y_{2[1:i-1]}(t)\right);$$

$$h\left(Y_{2m_2}(t) \Big| Q(t), Y_{2[1:m_2-1]}(t)\right) = h\left(Y_{11}(t) \Big| Q(t), Y_{2[1:m_2-1]}(t)\right);$$

$$h\left(Y_{1k}(t) \Big| Q(t), Y_{2[1:m_2]}(t), Y_{1[1:k-1]}(t)\right) = h\left(Y_{1l}(t) \Big| Q(t), Y_{2[1:m_2]}(t), Y_{1[1:k-1]}(t)\right).$$

Proof: It is sufficient to prove the first equality. Define $Y_2'(t) = Y_2(t) - H_{22}(t)X_2(t) = H_{21}(t)X_1(t) + W_2(t)$.

Toward this end, we have the following sequence of equalities,

$$h\Big(Y_{2i}(t)\Big|Q(t), Y_{2[1:i-1]}(t)\Big)$$

$$= h\Big(Y_{2i}(t)\Big|\mathcal{M}_2, \overline{H}(t), \overline{Y}_2(t-1), \overline{X}_2(t), Y_{2[1:i-1]}(t)\Big) \tag{14}$$

$$= h\Big(Y'_{2i}(t)\Big|\mathcal{M}_2, \overline{H}(t-1), \overline{Y}_2(t-1), \overline{X}_2(t), Y'_{2[1:i-1]}(t), H_{2[1:i]1}(t), H_{11}(t), H_{12}(t), H_{22}(t), H_{2[i+1:N_2]1}(t)\Big) \tag{15}$$

$$= h\Big(Y'_{2i}(t)\Big|\mathcal{M}_2, \overline{H}(t-1), \overline{Y}_2(t-1), \overline{X}_2(t), Y'_{2[1:i-1]}(t), H_{2[1:i]1}(t)\Big) \tag{16}$$

$$= \mathbb{E}_{H_{2i1}(t)=a}\, h\Big(Y'_{2i}(t)\Big|\mathcal{M}_2, \overline{H}(t-1), \overline{Y}_2(t-1), \overline{X}_2(t), Y'_{2[1:i-1]}(t), H_{2[1:i-1]1}(t), H_{2i1}(t)=a\Big) \tag{17}$$

$$= \mathbb{E}_{H_{2j1}(t)=a}\, h\Big(Y'_{2j}(t)\Big|\mathcal{M}_2, \overline{H}(t-1), \overline{Y}_2(t-1), \overline{X}_2(t), Y'_{2[1:i-1]}(t), H_{2[1:i-1]1}(t), H_{2j1}(t)=a\Big) \tag{18}$$

$$= h\Big(Y'_{2j}(t)\Big|\mathcal{M}_2, \overline{H}(t-1), \overline{Y}_2(t-1), \overline{X}_2(t), Y'_{2[1:i-1]}(t), H_{2[1:i-1]1}(t), H_{2j1}(t)\Big) \tag{19}$$

$$= h\Big(Y_{2j}(t)\Big|\mathcal{M}_2, \overline{H}(t-1), \overline{Y}_2(t-1), \overline{X}_2(t), Y_{2[1:i-1]}(t), H(t)\Big), \tag{20}$$

where the various equalities hold as follows: (14) holds by the definition of $Q(t)$; (15) holds because translation does not change differential entropy [1]; (16) follows by noting that $\{H_{11}(t), H_{12}(t), H_{22}(t), H_{2[i+1:N_2]1}(t)\}$ are independent of $\{Y'_{2i}(t), \mathcal{M}_2, \overline{H}(t-1), \overline{Y}_2(t-1), \overline{X}_2(t), Y'_{2[1:i-1]}(t), H_{2[1:i]1}(t)\}$ (note the present channel matrices are independent of the present and the past channel inputs and noises); (17) holds by the definition of the conditional differential entropy; (18) holds because conditioned on $\{\mathcal{M}_2, \overline{H}(t-1), \overline{Y}_2(t-1), \overline{X}_2(t), Y'_{2[1:i-1]}(t), H_{2[1:i-1]1}(t)\}$, the joint distribution of $\{H_{2i1}(t), X_1(t), W_{2i}(t)\}$ is identical to that of $\{H_{2j1}(t), X_1(t), W_{2j}(t)\}$; (19) holds by the definition of the conditional differential entropy; (20) holds since $\{H_{2i1}(t), H_{2[i+2:N_2]1}(t), H_{11}(t), H_{12}(t), H_{22}(t)\}$ are independent of $\{Y'_{2j}(t), \mathcal{M}_2, \overline{H}(t-1), \overline{Y}_2(t-1), \overline{X}_2(t), Y'_{2[1:i-1]}(t), H_{2[1:i-1]1}(t), H_{2j1}(t)\}$ and since translation does not change differential entropy. ∎

Note that the first equality in the above lemma asserts that the signals $Y_{2i}(t)$ and $Y_{2j}(t)$ received at the ith and j^{th} antenna, respectively, of R2 have equal differential entropy, when conditioned on the channel matrices $\overline{H}(t)$, the message $\mathcal{M}_2$ and the transmit signal $\overline{X}_2(t)$ of T2, the past channel outputs $\overline{Y}_2(t-1)$, and the present channel outputs $Y_{2[1:i-1]}(t)$ at some other receive antennas. We refer to this property as the statistical equivalence of the channel outputs, which essentially says that given the past and present channel outputs, the signals received at any two antennas of the system provide an equal amount of information about $\mathcal{M}_1$. Note that this property of statistical equivalence of the channel outputs was shown to hold in [20] for the case of delayed CSIT. Here, on the other hand, the same property is shown to be true under the stronger setting of Shannon feedback.

The above lemma yields the following simple corollary, where $Q(t) = \{\mathcal{M}_2, \overline{H}(t), \overline{Y}_2(t-1), \overline{X}_2(t)\}$ as before.

Corollary 3: For an $i \in [1:m_2-1]$ and a $k \in [1:m_1-m_2]$, if $j = i+1$ and $l = k+1$, we have the following:

$$h\Big(Y_{2i}(t)\Big|Q(t), Y_{2[1:i-1]}(t)\Big) \geq h\Big(Y_{2j}(t)\Big|Q(t), Y_{2[1:i]}(t)\Big);$$

$$h\Big(Y_{2m_2}(t)\Big|Q(t), Y_{2[1:m_2-1]}(t)\Big) \geq h\Big(Y_{11}(t)\Big|Q(t), Y_{2[1:m_2]}(t)\Big);$$

$$h\Big(Y_{1k}(t)\Big|Q(t), Y_{2[1:m_2]}(t), Y_{1[1:k-1]}(t)\Big) \geq h\Big(Y_{1l}(t)\Big|Q(t), Y_{2[1:m_2]}(t), Y_{1[1:k]}(t)\Big).$$

Proof: Follows from the previous lemma by invoking the fact that conditioning reduces entropy [1]. ∎

Lemma 6: We have

$$m_1 \cdot h\Big(Y_{2[1:m_2]}(t)\Big|Q(t)\Big) \geq m_2 \cdot h\Big(Y_{1[1:m_1-m_2]}(t), Y_{2[1:m_2]}(t)\Big|Q(t), \overline{Y}_1(t-1)\Big).$$

Proof: By the previous corollary and the chain rule for the differential entropy, we get

$$\frac{1}{m_2}h\Big(Y_{2[1:m_2]}(t)\Big|Q(t)\Big) = \frac{1}{m_2}\sum_{i=1}^{m_2}h\Big(Y_{2i}(t)\Big|Q(t), Y_{2[1:i-1]}(t)\Big)$$

$$\geq h\Big(Y_{2m_2}(t)\Big|Q(t), Y_{2[1:m_2-1]}(t)\Big)$$

$$\geq h\Big(Y_{11}(t)\Big|Q(t), Y_{2[1:m_2]}(t)\Big)$$

$$\geq \frac{1}{m_1-m_2}h\Big(Y_{1[1:m_1-m_2]}(t)\Big|Q(t), Y_{2[1:m_2]}(t)\Big).$$

This yields

$$(m_1-m_2)\cdot h\Big(Y_{2[1:m_2]}(t)\Big|Q(t)\Big) \geq m_2\cdot h\Big(Y_{1[1:m_1-m_2]}(t)\Big|Q(t), Y_{2[1:m_2]}(t)\Big)$$

$$\geq m_2\cdot h\Big(Y_{1[1:m_1-m_2]}(t)\Big|Q(t), \overline{Y}_1(t-1), Y_{2[1:m_2]}(t)\Big) \qquad (21)$$

since conditioning reduces entropy. Similarly, we can obtain

$$m_2\cdot h\Big(Y_{2[1:m_2]}(t)\Big|Q(t)\Big) \geq m_2\cdot h\Big(Y_{2[1:m_2]}(t)\Big|Q(t), \overline{Y}_1(t-1)\Big). \qquad (22)$$

The lemma can now be obtained by adding the inequalities in (21) and (22). ∎

The inequality in Lemma 2 can now be derived by combining the results of Lemmas 3, 4, and 6, and by noting that the sum or the difference of two $o(\log_2 P)$ terms yields another $o(\log_2 P)$ term.

IV. PROOF OF THEOREM 2

As mentioned before, it is sufficient to prove that the outer-bound $\mathbf{D}_{\text{outer}}^{\text{S}}$ is achievable when Condition 1 holds. Throughout this section, it is assumed that Condition 1 holds.

Here, bound L_2 can be easily shown to be redundant (it is implied by L_3), and thus can be ignored. Further, in the present case, bounds L_1 and L_3 are given by

$$L_1 \equiv \frac{d_1}{M_1'} + \frac{d_2}{N_2} \leq 1 \quad \text{and} \quad L_3 \equiv d_1 + d_2 \leq N_1,$$

where $M_1' \triangleq \min(M_1, N_1 + N_2)$.

The typical shape of the outer-bound is shown in Fig. 9, where $P_{o2,1}$ is the point of intersection of the line $d_2 = M_2$ and the one corresponding to bound L_1, similarly $P_{o2,3}$, and $P_{1,3}$ is the point of intersection of lines corresponding to bounds L_1 and L_3. Moreover,

$$P_{02,1} \equiv \left(M_1'\frac{N_2-M_2}{N_2}, M_2\right), \; P_{1,3} \equiv \left(M_1'\frac{N_1-N_2}{M_1'-N_2}, N_2\frac{M_1'-N_1}{M_1'-N_2}\right), \; \text{and} \; P_{o2,3} \equiv (N_1-M_2, M_2).$$

Depending on whether the d_2-coordinate of $P_{1,3}$ is less than M_2 or not, we have to consider two cases separately.

• Case A: $M_1' \geq N_2\frac{N_1-M_2}{N_2-M_2}$:

Here, bound L_1 is redundant. Moreover, from Fig. 9, one may observe that if $P_{o2,3} \in \mathbf{D}^{\text{S}}$ then $\mathbf{D}_{\text{outer}}^{\text{S}} = \mathbf{D}^{\text{S}}$. Hence, we find here sufficient to prove that $P_{o2,3} \in \mathbf{D}^{\text{S}}$.

• Case B: $M_1' < N_2\frac{N_1-M_2}{N_2-M_2}$:

Here, bounds L_1 and L_3 are both active. From Fig. 9, we observe the sufficiency of proving that $P_{o2,1}, P_{1,3} \in \mathbf{D}^{\text{S}}$.

Next, we propose a generic retrospective interference alignment scheme, which is used later to prove that $P_{o2,3} \in \mathbf{D}^{\text{S}}$ under Case A and $P_{o2,1}, P_{1,3} \in \mathbf{D}^{\text{S}}$ under Case B with an appropriate choice of parameters. This scheme is specified in terms of the parameters

$$T, \; t_1, \; t_2, \; \Big\{m_1(i)\Big\}_{i=1}^{T}, \; \text{and} \; \Big\{m_2(i)\Big\}_{i=1}^{T}, \qquad (23)$$

where $T, t_1, t_2 \in \mathbb{N}$, $m_1(i), m_2(i) \in \mathbb{N} \cup \{0\} \; \forall \; i$, and Design Criteria 1-5, which are stated in the sequel. It is developed such that if, for a given a DoF pair $P \equiv (d_1, d_2)$ and the given (M_1, M_2, N_1, N_2) MIMO IC, the
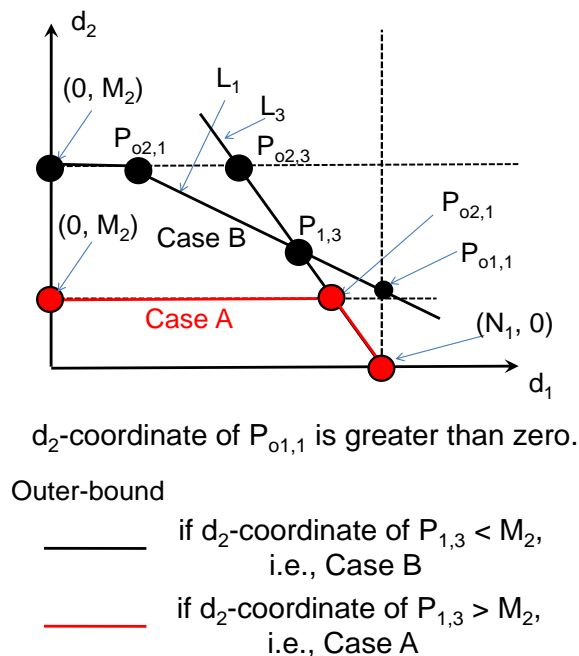
Fig. 9. Two Possible Shapes of the Outer-Bound when Condition 1 Holds

parameters in (23) can be chosen as functions of (d_1, d_2) and (M_1, M_2, N_1, N_2) so that Design Criteria 1-5 are satisfied, then the DoF pair $P \equiv (d_1, d_2) \in \mathbf{D}^{\mathrm{S}}$ of the given (M_1, M_2, N_1, N_2) MIMO IC.

Consider now the retrospective interference alignment scheme. The goal is to prove that a given DoF pair $P \equiv (d_1, d_2) \in \mathbf{D}^{\mathrm{S}}$. Let us first state two important design criteria.

Design Criterion 1: Choose positive integers t_1 and t_2 such that $t_1 + t_2 = T$.

Design Criterion 2: Choose a positive integer T such that Td_1 and Td_2 are integers.

Now, choose a $B \in \mathbb{N}$ and set

$$ T^\star = (B+1) \cdot T, \ d_i^\star = B \cdot T \cdot d_i, \text{ where } i \in \{1,2\}, \text{ and } P^\star \equiv \left(\frac{d_1^\star}{T^\star}, \frac{d_1^\star}{T^\star} \right) = \left(\frac{d_1}{1 + \frac{1}{B}}, \frac{d_2}{1 + \frac{1}{B}} \right). $$

It will be proved that for any positive integer B, by coding over $T^\star$ time slots, we can simultaneously achieve $d_1^\star$ and $d_2^\star$ DoF for the two users respectively. This implies that $P \in \mathbf{D}^{\mathrm{S}}$, since the DoF region is closed, and the point $P^\star$ converges to P as $B \to \infty$. Thus, our aim in the following is to prove the achievability of point $P^\star$.

The entire duration of $T^\star$ is divided into $B + 1$ blocks, each consisting of T time slots. Each block is further divided into two phases with Phase One having t_1 time slots and Phase Two the remainder of $t_2 = T - t_1$ time slots.

Definition 3: Define two functions $b(t)$, the index of the block to which time slot t belongs, and $\bar{t}(t)$, the index of that time slot within Block $b(t)$, as

$$ b(t) = \left\lceil \frac{t}{T} \right\rceil \quad \text{and} \quad \bar{t}(t) = t - T \cdot (b(t) - 1). $$

Note that $\bar{t}(t) \in [1 : T]$. Thus, each time slot t can be uniquely identified by the pair $(b(t), \bar{t}(t))$. Block b, $b \in [1 : B+1]$, consists of time slots $t \in [b'T + 1 : b'T + T]$, where $b' \triangleq b - 1$; we let Phase One of block b consist of time slots $t \in [b'T + 1 : b'T + t_1]$ and Phase Two the remaining time slots $t \in [b'T + t_1 + 1 : b'T + t_1 + t_2]$.

The general structure of our achievability scheme has the following features:

- In each of the first B blocks, T1 and T2 respectively transmit Td_1 and Td_2 i.i.d. complex Gaussian data symbols (DSs) intended for R1 and R2, respectively. In Block $B + 1$, no new DS is sent.
- In each time slot, T2 transmits an appropriate number of new DSs intended for R2, and therefore, in some sense, it does not play an active role in aligning interference (as in the example of Section II-C).

- T1, on the other hand, transmits all Td_1 DSs, to be transmitted over a given block during Phase One of that block (time slots $t = 1$ or $t = 4$ constitute Phase One in the example of Section II-C). It signals over Phase Two such that at the end of each block, (a) R2 can decode all DSs sent to it over that block, and (b) R1 can decode all DSs sent to it over the previous block (note that $t = 2, 3$ or $t = 4, 5$ are counterparts of Phase Two in the example of Section II-C). To meet these objectives optimally, T1 needs to align interference at both receivers.

- Finally, the goal of Block $B + 1$ is to allow R1 to decode all DSs of Block B by not sending any new data (Block $B + 1$, in some sense, corresponds to time slot $t = 7$ in the example of Section II-C).

Let T1 and T2 transmit $m_1(i)$ and $m_2(i)$ DSs intended for their respective receivers at the ith time slot of any given block (except, the last block) where these design parameters are chosen according to following criterion.

Design Criterion 3: Choose non-negative integers $m_1(i)$ and $m_2(i)$, $i \in [1 : T]$, as follows.

- $m_1(i) = 0 \ \forall \ i \in [t_1 + 1 : T]$ (recall, $t_1 + t_2 = T$);
- $m_1(i) \leq M_1' \ \forall \ i \in [1 : t_1]$ and $\sum_{i=1}^{T} m_1(i) = \sum_{i=1}^{t_1} m_1(i) = Td_1$;
- $m_2(j) \leq M_2 \ \forall \ j \in [1 : T]$ and $\sum_{j=1}^{T} m_2(j) = Td_2$.

At time $t \in [1 : BT]$, T1 transmits $m_1\big(\bar{t}(t)\big)$ complex Gaussian DSs, denoted by $u_{1i}\big(b(t), \bar{t}(t)\big)$, whereas T2 transmits $m_2\big(\bar{t}(t)\big)$ complex Gaussian DSs, denoted by $u_{2i}\big(b(t), \bar{t}(t)\big)$. Note that all DSs

$$\left\{ \left\{ \{u_{1i}(b,\bar{t})\}_{i=1}^{m_1(\bar{t})} \right\}_{\bar{t}=1}^{T} \right\}_{b=1}^{B} \quad \text{and} \quad \left\{ \left\{ \{u_{2j}(b,\bar{t})\}_{j=1}^{m_2(\bar{t})} \right\}_{\bar{t}=1}^{T} \right\}_{b=1}^{B}$$

are i.i.d. Note that since $m_1(i) = 0 \ \forall \ i \in [t_1 + 1 : t_1 + t_2]$, no new DS is transmitted by T1 to R1 over Phase Two of any block so as to enable interference alignment and to ensure successful decoding.

The transmission scheme of T1 and T2 are described next. Focusing on Block b, where $b \in [1 : B]$, the operation of T1 and T2 is described over the two phases separately. We start below with Phase One of Block b. See also Tables I-III, where the main points about the operation of this scheme are summarized.

<u>Block b, Phase One:</u> Here, $t \in [b'T + 1 : b'T + t_1]$ with $b \in [1 : B]$ and $b' = b - 1$. This phase is a data transmission phase.

At time $t \in [b'T + 1 : b'T + t_1]$ with $b \in [1 : B]$, T1 and T2 respectively transmit $m_1\big(\bar{t}(t)\big)$ and $m_2\big(\bar{t}(t)\big)$ DSs as follows:

$$
\begin{aligned}
X_{1i}(t) &= u_{1i}\big(b(t), \bar{t}(t)\big), \ \forall \, i \in \left[1 : m_1\big(\bar{t}(t)\big)\right], \ \text{and} \\
X_{1i}(t) &= 0, \ \forall \, i \in \left[m_1\big(\bar{t}(t)\big) + 1 : M_1\right]; \ \text{and} \\
X_{2j}(t) &= u_{2j}\big(b(t), \bar{t}(t)\big), \ \forall \, j \in \left[1 : m_2\big(\bar{t}(t)\big)\right], \ \text{and} \\
X_{2j}(t) &= 0, \ \forall \, j \in \left[m_2\big(\bar{t}(t)\big) + 1 : M_2\right].
\end{aligned}
$$

Consider now the signals received by R1 and R2 during Phase One. Since we are interested in the achievability of the DoF, we ignore throughout the presence of additive noise since it can not affect a DoF result. Then for a

$t \in [b'T + 1 : b'T + t_1]$ with $b \in [1 : B]$, we have

$$Y_{1i}(t) = H_{1i1}(t)X_1(t) + H_{1i2}(t)X_2(t) \cdots i \in [1 : N_1]$$

$$= H_{1i1}(t) \underbrace{\begin{bmatrix} u_{11}\big(b(t), \bar{t}(t)\big) \\ u_{12}\big(b(t), \bar{t}(t)\big) \\ \vdots \\ u_{1m_1(\bar{t}(t))}\big(b(t), \bar{t}(t)\big) \\ 0_{[M_1 - m_1(\bar{t}(t))] \times 1} \end{bmatrix}}_{\triangleq \, \mathrm{LC}_{1i}^{[1]}\big(b(t), \bar{t}(t)\big)} + H_{1i2}(t) \underbrace{\begin{bmatrix} u_{21}\big(b(t), \bar{t}(t)\big) \\ u_{22}\big(b(t), \bar{t}(t)\big) \\ \vdots \\ u_{2m_2(\bar{t}(t))}\big(b(t), \bar{t}(t)\big) \\ 0_{[M_2 - m_2(\bar{t}(t))] \times 1} \end{bmatrix}}_{\triangleq \, \mathrm{LC}_{1i}^{[2]}\big(b(t), \bar{t}(t)\big)};$$

$$Y_{2j}(t) = H_{2j1}(t)X_1(t) + H_{2j2}(t)X_2(t) \cdots j \in [1 : N_2]$$

$$= H_{2j1}(t) \underbrace{\begin{bmatrix} u_{11}\big(b(t), \bar{t}(t)\big) \\ u_{12}\big(b(t), \bar{t}(t)\big) \\ \vdots \\ u_{1m_1(\bar{t}(t))}\big(b(t), \bar{t}(t)\big) \\ 0_{[M_1 - m_1(\bar{t}(t))] \times 1} \end{bmatrix}}_{\triangleq \, \mathrm{LC}_{2j}^{[1]}\big(b(t), \bar{t}(t)\big)} + H_{2j2}(t) \underbrace{\begin{bmatrix} u_{21}\big(b(t), \bar{t}(t)\big) \\ u_{22}\big(b(t), \bar{t}(t)\big) \\ \vdots \\ u_{2m_2(\bar{t}(t))}\big(b(t), \bar{t}(t)\big) \\ 0_{[M_2 - m_2(\bar{t}(t))] \times 1} \end{bmatrix}}_{\triangleq \, \mathrm{LC}_{2j}^{[2]}\big(b(t), \bar{t}(t)\big)}.$$

Here, $\mathrm{LC}_{ij}^{[k]}(b, \bar{t})$ represents the linear combination of DSs sent by the kth transmitter at time $t = (b-1)T + \bar{t}$, and it affects the signal received by the ith receiver at its jth antenna. Note here that $\mathrm{LC}_{2j}^{[1]}(b(t), \bar{t}(t))$ is a linear combination of DSs intended only for R1, while it causes interference to R2. Note the collection

$$\Big\{ \big\{ \mathrm{LC}_{2j}^{[1]}(b, \bar{t}) \big\}_{j=1}^{N_2} \Big\}_{\bar{t}=1}^{t_1}$$

is referred to henceforth as the interference seen by R2 during Phase One of Block b. Evidently, interference at R2 is useful for R1. This completes the description of the operation over Phase One.

Consider next the second phase of Block $b \in [1 : B]$. Here, T2 continues to transmit data. T1, on the other hand, employs two interference alignment techniques (described in Section II-D) over Phase Two. In the first one, T1 transmits a part of the interference seen by R2 over Phase One of the previous block (this idea is not used during Block 1), and under the second technique, it transmits all DSs of T2 sent earlier over Phase One of the same block. Now, as shown shortly, if over each time slot of Phase Two T1 transmits an appropriate numbers of DSs of T2 and the interfering linear combinations at R2 then at the end of each block R2 can decode the desired DSs sent over the same block, while R1 can decode those sent over the previous block.

Lemma 7: At time $t = b'T + t_1 + 1$, T1 can obtain (a) DSs $\Big\{ u_{2i}\big(b(t), \bar{t}(t)\big) \Big\} \, \forall \, i \in \Big[1 : m_2\big(\bar{t}(t)\big)\Big]$ and $t \in \Big[b'T + 1 : b'T + t_1\Big]$, and (b) linear combinations $\Big\{ \mathrm{LC}_{2j}^{[1]}\big(b(t), \bar{t}(t)\big) \Big\}, \forall \, j \in [1 : N_2]$ and $t \in \Big[b'T + 1 : b'T + t_1\Big]$.

Proof: Consider symbols in (a). By virtue of Shannon feedback, T1 knows all past channel matrices $H(t)$, the received signal $Y_2(t)$, as well as its own transmit signal. Hence, for each $t \in [b'T + 1 : b'T + t_1]$, it can compute

$$Y_2(t) - H_{21}(t)X_1(t) = H_{22}(t)X_2(t).$$

Since $N_2 > M_2$ under Condition 1, and the channel matrices are Rayleigh faded, $H_{22}(t)$ is almost surely a one-to-one map. Hence, based on the value of $H_{22}(t)X_2(t)$, T1 can determine $X_2(t)$. In other words, at $t = b'T + t_1 + 1$, T1 can perfectly evaluate data symbols

$$\Big\{ \big\{ u_{2i}\big(b(t), \bar{t}\big) \big\}_{i=1}^{m_2(\bar{t})} \Big\}_{\bar{t}=1}^{t_1}.$$

That T1 can obtain the symbols in part (b) of the lemma can be shown analogously. ∎

Next, consider the transmission strategy of T1 over Phase Two. We need to construct two sets $\big\{ \mathcal{P}_{\mathrm{LC}}(b, i) \big\}_{i=1}^{t_2}$ and $\big\{ \mathcal{P}_{\mathrm{DS}}(b, i) \big\}_{i=1}^{t_2}$ for each $b \in [1 : B+1]$. The set $\mathcal{P}_{\mathrm{LC}}(b, i)$ contains the linear combinations that interfere with R2 over Phase One of Block $b-1$ (except, for $b = 1$, for which this set is empty) and that are to be sent by T1 over the ith time slot of Phase Two of Block b. Moreover, the set $\mathcal{P}_{\mathrm{DS}}(b, i)$ contains DSs that are sent by T2 over

Phase One, Block 1: t is such that $b(t) = 1$ and $\bar{t}(t) \in [1 : t_1]$.

node	operation at time t
T1	transmits DSs $u_{1i}(1, \bar{t}(t))$, $i \in [1 : m_1(\bar{t}(t))]$.
T2	transmits DSs $u_{2j}(1, \bar{t}(t))$, $j \in [1 : m_2(\bar{t}(t))]$.
R1	receives $Y_{1i}(t) = \mathrm{LC}_{1i}^{[1]}(1, \bar{t}(t)) + \mathrm{LC}_{1i}^{[2]}(1, \bar{t}(t))$, $i \in [1 : N_1]$.
R2	receives $Y_{2j}(t) = \mathrm{LC}_{2j}^{[1]}(1, \bar{t}(t)) + \mathrm{LC}_{2j}^{[2]}(1, \bar{t}(t))$, $j \in [1 : N_2]$.

Phase Two, Block 1: t is such that $b(t) = 1$ and $\bar{t}(t) \in [1 : t_2]$.

node	operation
T1	retransmits DSs $\left\{ \left\{ u_{2i}(1, \bar{t}) \right\}_{i=1}^{m_2(\bar{t})} \right\}_{\bar{t}=1}^{t_1}$ sent by T2 over Phase One. At any time at most $N_2 - M_2$ antennas are active.
T2	transmits $u_{2j}(1, \bar{t}(t))$, $j \in [1 : m_2(\bar{t}(t))]$.
R1	Gets linear combinations $\left\{ \left\{ \mathrm{LC}_{1i}^{[1]}(1, \bar{t}) \right\}_{i=1}^{N_1} \right\}_{\bar{t}=1}^{t_1}$.
R2	Can decode all DSs sent over this block. Knows the values of linear combinations $\left\{ \left\{ \mathrm{LC}_{2j}^{[1]}(1, \bar{t}) \right\}_{j=1}^{N_2} \right\}_{\bar{t}=1}^{t_1}$.

TABLE I

BLOCK 1 OF THE RETROSPECTIVE INTERFERENCE ALIGNMENT SCHEME WITH SHANNON FEEDBACK

Phase One of Block b and that are to be retransmitted by T1 over the ith time slot of Phase Two of Block b. In terms of these sets the transmissions of T1 and T2 are then described.

With an aim of constructing sets $\left\{ \mathcal{P}_{\mathrm{LC}}(b, i) \right\}_{i=1}^{t_2}$ and $\left\{ \mathcal{P}_{\mathrm{DS}}(b, i) \right\}_{i=1}^{t_2}$ for each $b \in [1 : B + 1]$, consider the following. Let $(c - d)^+ \triangleq \max\{0, c - d\}$, $|\mathcal{S}|$ denote the cardinality of set $\mathcal{S}$, $\mathcal{S}_{\mathrm{LC}}(1) = \phi$ (the empty set), and $\mathcal{S}_{\mathrm{DS}}(B + 1) = \phi$. Further, define

$$n_{\mathrm{req}}(i) \triangleq \left(m_1(i) - N_1 \right)^+, \text{ where } i \in [1 : t_1]$$

$$\text{for a } b \in [2 : B + 1], \mathcal{S}_{\mathrm{LC}}(b) \triangleq \left\{ \left\{ \mathrm{LC}_{2i}^{[1]}\left(b - 1, \bar{t} \right) \right\}_{i=1}^{n_{\mathrm{req}}(\bar{t})} \right\}_{\bar{t}=1}^{t_1}, \text{ and}$$

$$\text{for a } b \in [1 : B], \mathcal{S}_{\mathrm{DS}}(b) \triangleq \left\{ \left\{ u_{2i}\left(b, \bar{t} \right) \right\}_{i=1}^{m_2(\bar{t})} \right\}_{\bar{t}=1}^{t_1}.$$

Here, the set $\mathcal{S}_{\mathrm{LC}}(b)$, $b \geq 2$, contains linear combinations that interfere with R2 over Block $b - 1$ and the elements of this set are chosen such that if all linear combinations in this set are delivered to R1 then R1 can decode all desired DSs sent over Block $b - 1$. Moreover, sets $\left\{ \mathcal{P}_{\mathrm{LC}}(b, i) \right\}_{i=1}^{t_2}$ are constructed by partitioning the set $\mathcal{S}_{\mathrm{LC}}(b)$. Further, the set $\mathcal{S}_{\mathrm{DS}}(b)$, $b \leq B$, contains all DSs sent by T2 over Phase One of Block b, and by partitioning this set, smaller sets $\left\{ \mathcal{P}_{\mathrm{DS}}(b, i) \right\}_{i=1}^{t_2}$ are formed.

Note that

$$\left| \mathcal{S}_{\mathrm{LC}}(b) \right| = \sum_{\bar{t}=1}^{t_1} n_{\mathrm{req}}(\bar{t}) \text{ and } \left| \mathcal{S}_{\mathrm{DS}}(b) \right| = \sum_{\bar{t}=1}^{t_1} m_2(\bar{t}).$$

Consider next two more design criteria which ensure that the cardinalities of these sets are appropriately bounded.

Design Criterion 4: Choose t_1, t_2, and $m_2(\bar{t})$, where $\bar{t} \in [1 : t_1]$, such that

$$\left| \mathcal{S}_{\mathrm{DS}}(b) \right| = \sum_{\bar{t}=1}^{t_1} m_2(\bar{t}) \leq (N_2 - M_2) \cdot t_2.$$

Phase One, Block b, $b \in [2 : B]$:

node	operation at time t
T1	transmits DSs $u_{1i}(b, \bar{t}(t))$, $i \in [1 : m_1(\bar{t}(t))]$.
T2	transmits DSs $u_{2j}(b, \bar{t}(t))$, $j \in [1 : m_2(\bar{t}(t))]$.
R1	receives $Y_{1i}(t) = \mathrm{LC}_{1i}^{[1]}(b, \bar{t}(t)) + \mathrm{LC}_{1i}^{[2]}(b, \bar{t}(t))$, $i \in [1 : N_1]$.
R2	receives $Y_{2j}(t) = \mathrm{LC}_{2j}^{[1]}(b, \bar{t}(t)) + \mathrm{LC}_{2j}^{[2]}(b, \bar{t}(t))$, $j \in [1 : N_2]$.

Phase Two, Block b, $b \in [2 : B]$:

node	operation
T1	transmits DSs $\left\{ \{u_{2i}(b, \bar{t})\}_{i=1}^{m_2(\bar{t})} \right\}_{\bar{t}=1}^{t_1}$ and linear combinations $\left\{ \left\{ \mathrm{LC}_{2j}^{[1]}(b-1, \bar{t}) \right\}_{j=1}^{n_{\mathrm{req}}(\bar{t})} \right\}_{\bar{t}=1}^{t_1}$. At any time, at most $N_2 - M_2$ DSs and $N_1 - N_2$ linear combinations are transmitted.
T2	transmits $u_{2j}(b, \bar{t}(t))$, $j \in [1 : m_2(\bar{t}(t))]$.
R1	Gets linear combinations $\left\{ \left\{ \mathrm{LC}_{1i}^{[1]}(b, \bar{t}) \right\}_{i=1}^{N_1} \right\}_{\bar{t}=1}^{t_1}$ and $\left\{ \left\{ \mathrm{LC}_{2j}^{[1]}(b-1, \bar{t}) \right\}_{j=1}^{n_{\mathrm{req}}(\bar{t})} \right\}_{\bar{t}=1}^{t_1}$. Can decode all DSs sent over Block $b-1$.
R2	Can decode all DSs sent over this Block. Knows the values of linear combinations $\left\{ \left\{ \mathrm{LC}_{2j}^{[1]}(b, \bar{t}) \right\}_{j=1}^{N_2} \right\}_{\bar{t}=1}^{t_1}$.

TABLE II

BLOCK b, $b \in [2 : B]$, OF RETROSPECTIVE IA SCHEME WITH SHANNON FEEDBACK

Design Criterion 5: Choose t_1, t_2, and $m_1(\bar{t})$, where $\bar{t} \in [1 : t_1]$, such that

$$\left| \mathcal{S}_{\mathrm{LC}}(b) \right| = \sum_{\bar{t}=1}^{t_1} n_{\mathrm{req}}(\bar{t}) \leq (N_1 - N_2) \cdot t_2.$$

If these two criteria are satisfied, sets $\mathcal{S}_{\mathrm{DS}}(b)$ and $\mathcal{S}_{\mathrm{LC}}(b)$ can be partitioned as follows.

- Partition $\mathcal{S}_{\mathrm{LC}}(b)$ into t_2 disjoint subsets each of cardinality at most $N_1 - N_2$ so that

$$\mathcal{S}_{\mathrm{LC}}(b) = \bigcup_{i=1}^{t_2} \mathcal{P}_{\mathrm{LC}}(b, i) \text{ and } \left| \mathcal{P}_{\mathrm{LC}}(b, i) \right| \leq N_1 - N_2. \tag{24}$$

- Partition $\mathcal{S}_{\mathrm{DS}}(b)$ into t_2 disjoint subsets each of cardinality at most $N_2 - M_2$ so that

$$\mathcal{S}_{\mathrm{DS}}(b) = \bigcup_{i=1}^{t_2} \mathcal{P}_{\mathrm{DS}}(b, i) \text{ and } \left| \mathcal{P}_{\mathrm{DS}}(b, i) \right| \leq N_2 - M_2. \tag{25}$$

Suppose for each $i \in [1 : t_2]$,

$$\mathcal{P}_{\mathrm{LC}}(b, i) \bigcup \mathcal{P}_{\mathrm{DS}}(b, i) = \left\{ p_j(b, i) \right\}, \ j \in \left[1 : \left| \mathcal{P}_{\mathrm{LC}}(b, i) \right| + \left| \mathcal{P}_{\mathrm{DS}}(b, i) \right| \right].$$

Over Phase Two of Block b, T1 transmits all the elements of the set $\mathcal{S}_{\mathrm{LC}}(b) \bigcup \mathcal{S}_{\mathrm{DS}}(b)$, while T2 continues to transmit DSs intended for R2. See also Tables I-III.

Block b, Phase Two: Here, $t \in [b'T + t_1 + 1 : b'T + t_1 + t_2]$.

At time $t \in [b'T + t_1 + 1 : b'T + t_1 + t_2]$ with $b \in [1 : B + 1]$, T1 transmits the elements of set $\mathcal{P}_{\mathrm{LC}}\left(b, \bar{t}(t) - t_1\right) \bigcup \mathcal{P}_{\mathrm{DS}}\left(b, \bar{t}(t) - t_1\right)$ as follows:

$$X_{1i}(t) = \begin{cases} p_i\left(b(t), \bar{t}(t) - t_1\right), & \text{if } i \in \left[1 : \left| \mathcal{P}_{\mathrm{LC}}\left(b(t), \bar{t}(t) - t_1\right) \right| + \left| \mathcal{P}_{\mathrm{DS}}\left(b(t), \bar{t}(t) - t_1\right) \right| \right], \\ 0, & \text{if } i \in \left[1 + \left| \mathcal{P}_{\mathrm{LC}}\left(b(t), \bar{t}(t) - t_1\right) \right| + \left| \mathcal{P}_{\mathrm{DS}}\left(b(t), \bar{t}(t) - t_1\right) \right| : M_1 \right]. \end{cases}$$

During Phase Two of Block b, T2 transmits DSs $\left\{\left\{u_{2j}(b,\bar{t})\right\}_{j=1}^{m_2(\bar{t})}\right\}_{\bar{t}=t_1+1}^{T}$ if $b \leq B$, and remains silent over the last Block. Thus, we have the following: for $t \in [b'T + t_1 + 1 : b'T + t_1 + t_2]$ with $b \leq B$,

$$X_{2j}(t) = u_{2j}\Big(b(t), \bar{t}(t)\Big), \ j \in \Big[1 : m_2\big(\bar{t}(t)\big)\Big], \text{ and}$$

$$X_{2j}(t) = 0, \ j \in \Big[m_2\big(\bar{t}(t)\big) + 1 : M_2\Big],$$

whereas for $t \in [BT + t_1 + 1 : BT + t_1 + t_2]$ (i.e., over Block $B + 1$),

$$X_2(t) = 0.$$

This completes the description of the transmission strategy of T1 and T2.

Consider now the decoding operation starting with R2. The following two lemmas enable an inductive proof that R2 can decode all desired DSs. See also Tables I-III.

The next lemma proves that R2 can decode all DSs sent to it over Block 1.

Lemma 8: At time $t = T$, R2 can decode DSs

$$\left\{\left\{u_{2j}(1,\bar{t})\right\}_{j=1}^{m_2(\bar{t})}\right\}_{\bar{t}=1}^{T}$$

sent to it over Block 1.

Proof: Recall that for Block 1, the set $\mathcal{S}_{\mathrm{LC}}(1)$ is empty. Thus, over Phase Two of Block 1, T1 transmits all elements of the set $\mathcal{S}_{\mathrm{DS}}(1)$, i.e., T1 retransmits all DSs $\left\{\left\{u_{2i}(1,\bar{t})\right\}_{i=1}^{m_2(\bar{t})}\right\}_{\bar{t}=1}^{t_1}$ that are sent by T2 over Phase One of this block. Moreover, since $|\mathcal{P}_{\mathrm{DS}}(1,i)| \leq N_2 - M_2, \ \forall \ i, b$, T1, at any time during Phase Two, transmits at most $N_2 - M_2$ elements of the set $\mathcal{S}_{\mathrm{DS}}(1) = \left\{\left\{u_{2i}(1,\bar{t})\right\}_{i=1}^{m_2(\bar{t})}\right\}_{\bar{t}=1}^{t_1}$. This implies that at any time during Phase Two of Block 1, at most N_2 transmit antennas are active (i.e., they send a non-zero signal). Since the Rayleigh-faded channel matrices are full rank almost surely, R2, equipped with N_2 antennas, can determine the transmit signal via simple channel inversion. Therefore, at time $t \in [t_1 + 1 : T]$, R2 can decode all DSs belonging to the set $\mathcal{P}_{\mathrm{DS}}\Big(b(t), \bar{t}(t) - t_1\Big)$ and also those transmitted by T2 at that time, namely, $u_{2i}\Big(1, \bar{t}(t)\Big), \ i \in \Big[1 : m_2\big(\bar{t}(t)\big)\Big]$. Therefore, at the end of Block 1, i.e., at $t = T$, R2 can decode all DSs sent to it over this block. ∎

The next lemma deals with decoding of DSs sent to R2 over Block b, $b \in [2 : B]$.

Lemma 9: Consider Block b, $b \in [2 : B]$. If, at time $t = (b-1)T$, R2 has successfully decoded all DSs

$$\left\{\left\{u_{2i}(b-1,\bar{t})\right\}_{i=1}^{m_2(\bar{t})}\right\}_{\bar{t}=1}^{T}$$

sent to it over Block $(b-1)$, then at time $t = bT$, it can decode all DSs

$$\left\{\left\{u_{2i}(b,\bar{t})\right\}_{i=1}^{m_2(\bar{t})}\right\}_{\bar{t}=1}^{T}$$

sent to it over Block b.

Proof: Suppose at time $t = (b-1)T$, R2 has decoded successfully all DSs sent to it over Block $(b-1)$. Thus, R2, at time $t = (b-1)T$, can determine the values of linear combinations

$$\mathrm{LC}_{2j}^{[2]}(b-1,\bar{t}) = H_{2j2}\big((b-2)T + \bar{t}\big) \begin{bmatrix} u_{21}\big(b-1,\bar{t}\big) \\ u_{22}\big(b-1,\bar{t}\big) \\ \vdots \\ u_{2m_2(\bar{t})}\big(b-1,\bar{t}\big) \\ 0_{[M_2 - m_2(\bar{t})] \times 1} \end{bmatrix}, \ \forall j \in [1 : N_2] \text{ and } \bar{t} \in [1 : t_1],$$

and hence, it can also evaluate

$$\mathrm{LC}_{2j}^{[1]}(b-1,\bar{t}) = Y_{2j}\big((b-2)T + \bar{t}\big) - \mathrm{LC}_{2j}^{[2]}(b-1,\bar{t}), \ \forall j \in [1 : N_2] \text{ and } \bar{t} \in [1 : t_1].$$

Phase One, Block $B + 1$: No operation is performed.

Phase Two, Block $B + 1$:

node	operation
T1	transmits linear combinations $\left\{ \left\{ \text{LC}_{2j}^{[1]}\left(B,\bar{t}\right) \right\}_{j=1}^{n_{\text{req}}(\bar{t})} \right\}_{\bar{t}=1}^{t_1}$.
T2	remains silent
R1	Gets linear combinations $\left\{ \left\{ \text{LC}_{2j}^{[1]}\left(B,\bar{t}\right) \right\}_{j=1}^{n_{\text{req}}(\bar{t})} \right\}_{\bar{t}=1}^{t_1}$. Can decode all DSs sent over Block B.
R2	idle

TABLE III

BLOCK $B + 1$ OF RETROSPECTIVE IA SCHEME WITH SHANNON FEEDBACK

In particular, R2 at time $t = (b-1)T$ knows values of all elements of set

$$\mathcal{S}_{\text{LC}}(b) = \left\{ \left\{ \text{LC}_{2i}^{[1]}\left(b-1,\bar{t}\right) \right\}_{i=1}^{n_{\text{req}}(\bar{t})} \right\}_{\bar{t}=1}^{t_1}.$$

Consider now the operation over Phase Two of Block b, i.e., for a $t \in [(b-1)T + t_1 + 1 : (b-1)T + t_1 + t_2]$. Since R2 already knows the values of elements of set $\mathcal{S}_{\text{LC}}(b)$, it can subtract from $Y_2(t)$ the contribution due to the elements of set $\mathcal{P}_{\text{LC}}\left(b, \bar{t}(t) - t_1\right) \subset \mathcal{S}_{\text{LC}}(b)$, which are transmitted by T1 at time t. After this subtraction, from the perspective of R2, not more than N_2 transmit antennas are active. Therefore, as mentioned before, R2 can use channel inversion to determine the elements of set $\mathcal{P}_{\text{DS}}\left(b, \bar{t}(t) - t_1\right)$ and also $u_{2i}\left(b, \bar{t}(t)\right) \ \forall \ i \in [1 : m_2(\bar{t}(t))]$. $\blacksquare$

Combined with the result of the previous lemma that all Block 1 DSs can be decided at $t = T$, the above lemma can now be applied recursively to show that R2 can decode all desired DSs sent over the first B blocks. Since no new DSs are transmitted over the last $(B+1)^{th}$ block, decoding is successful at R2.

Consider now the decoding procedure at R1. See also Tables I-III. It turns out that R1, at the end of a given block, does not observe a sufficient number of interference-free linear combinations required to decode desired DSs sent over that block. However, the missing linear combinations are sent to it over the next block. Hence, DSs sent over a given block are decodable at R1 at the end of the next block. Since no new DS is sent over the final block, decoding is successful at R1 at the end of Block $B + 1$.

The above claims about how decoding works at R1 are proved using a series of three lemmas. The first two specify the linear combinations that are known to R1 at the end of each block. The third lemma makes use of the first two to prove that all desired DSs are decodable at R1 at the end of Block $B + 1$.

Lemma 10: At the end of Block 1, i.e., time $t = T$, R1 can obtain linear combinations

$$\text{LC}_{1i}^{[1]}\left(1, \bar{t}(t)\right) \ \forall \ i \in [1 : N_1] \text{ and } t \in [1 : t_1].$$

Proof: As with R2, R1 at $t = T$ can determine the DSs

$$\left\{ \left\{ u_{2i}\left(1, \bar{t}(t)\right) \right\}_{i=1}^{m_2(\bar{t}(t))} \right\}_{t=1}^{t_1}.$$

Hence it can evaluate

$$\text{LC}_{1i}^{[2]}\left(1, \bar{t}\right) = H_{1i2}(\bar{t}) \begin{bmatrix} u_{21}\left(1,\bar{t}\right) \\ u_{22}\left(1,\bar{t}\right) \\ \vdots \\ u_{2m_2(\bar{t})}\left(1,\bar{t}\right) \\ 0_{\left[M_2 - m_2(\bar{t})\right] \times 1} \end{bmatrix}, \ \forall \ i \in [1 : N_1] \text{ and } \bar{t} \in [1 : t_1],$$

and then, $\text{LC}_{1i}^{[1]}\left(1, \bar{t}\right) = Y_{1i}(\bar{t}) - \text{LC}_{1i}^{[2]}\left(1, \bar{t}\right), \ \forall \ i \in [1 : N_1] \text{ and } \bar{t} \in [1 : t_1].$ $\blacksquare$

Lemma 11: At the end of Block b, $b \in [2 : B]$, i.e., at $t = bT$, R1 can obtain the linear combinations

$$\left\{ \{ \mathrm{LC}_{1i}^{[1]}(b, \bar{t}) \}_{i=1}^{N_1} \right\}_{\bar{t}=1}^{t_1} \text{ and } \left\{ \{ \mathrm{LC}_{2i}^{[1]}(b-1, \bar{t}) \}_{i=1}^{n_{\mathrm{req}}(\bar{t})} \right\}_{\bar{t}=1}^{t_1}.$$

Further, at the end of Block $B + 1$, R1 can obtain the linear combinations

$$\left\{ \{ \mathrm{LC}_{2i}^{[1]}(B, \bar{t}) \}_{i=1}^{n_{\mathrm{req}}(\bar{t})} \right\}_{\bar{t}=1}^{t_1}.$$

Proof: Consider the operation over Phase Two of Block b, $b \leq B$ i.e., for $t \in [b'T + t_1 + 1 : b'T + t_1 + t_2]$ with $b \leq B$. At any time during this phase, at most N_1 transmit antennas are active. Hence, via simple channel inversion, R1 can determine the transmit signal during this phase. Thus, at the end of this phase, i.e., at $t = bT$, R1 knows the values of elements of the sets $\mathcal{S}_{\mathrm{LC}}(b)$ and $\mathcal{S}_{\mathrm{DS}}(b)$. This implies that at the end of Block b R1 knows linear combinations $\left\{ \{ \mathrm{LC}_{2i}^{[1]}(b-1, \bar{t}) \}_{i=1}^{n_{\mathrm{req}}(\bar{t})} \right\}_{\bar{t}=1}^{t_1}$, which are contained in the set $\mathcal{S}_{\mathrm{LC}}(b)$. Further, it knows DSs $\left\{ \{ u_{2i}(b, \bar{t}) \}_{i=1}^{m_2(\bar{t})} \right\}_{\bar{t}=1}^{t_1}$, from which it can compute $\left\{ \{ \mathrm{LC}_{1i}^{[2]}(b, \bar{t}) \}_{i=1}^{N_1} \right\}_{\bar{t}=1}^{t_1}$, and subsequently obtain $\left\{ \{ \mathrm{LC}_{1i}^{[1]}(b, \bar{t}) \}_{i=1}^{N_1} \right\}_{\bar{t}=1}^{t_1}$.

The last part of the lemma can be proved using similar arguments. $\blacksquare$

Lemma 12: At the end of Block b, where $b \in [2 : B + 1]$, R1 can decode data symbols

$$\left\{ \{ u_{1i}(b-1, \bar{t}) \}_{i=1}^{m_1(\bar{t})} \right\}_{\bar{t}=1}^{t_1}$$

sent to it over Block $b - 1$.

Proof: First recall that no new DSs are transmitted to R1 during Block $B + 1$. Therefore, we will focus on Block $b - 1$, where $b \in [2 : B + 1]$. It will be shown that given any t such that $b(t) = b - 1$, all DSs sent by T1 to R1 at time t can be decoded by R1 at the end of Block b. Moreover, since no new DSs are sent to R1 over Phase Two of any block, we may assume, without loss of generality, that $\bar{t}(t) \in [1 : t_1]$. Thus, in the following, we consider a time slot t with $b(t) = (b - 1)$, $b \in [2 : B + 1]$, and $\bar{t}(t) \in [1 : t_1]$ or $t = (b - 2)T + \bar{t}$ with $\bar{t} \in [1 : t_1]$.

At this time, T1 transmits DSs

$$\left\{ u_{1i}\left(b - 1, \bar{t}(t) \right) \right\}, i \in \left[1 : m_1\left(\bar{t}(t) \right) \right].$$

To decode these DSs, it is sufficient for R1 to know $m_1\left(\bar{t}(t) \right)$ linearly independent linear combinations of these data symbols. Consider the following collection of linear combinations:

$$\left\{ \mathrm{LC}_{1i}^{[1]}\left(b - 1, \bar{t}(t) \right) \right\}_{i=1}^{N_1} \text{ and } \left\{ \mathrm{LC}_{2i}^{[1]}\left(b - 1, \bar{t}(t) \right) \right\}_{i=1}^{n_{\mathrm{req}}\left(\bar{t}(t) \right)}.$$

These are $m_1\left(\bar{t}(t) \right)$ linear combinations of DSs $\left\{ u_{1i}\left(b - 1, \bar{t}(t) \right) \right\}$, $i \in \left[1 : m_1\left(\bar{t}(t) \right) \right]$, and since the Rayleigh-faded channel matrices are full rank with probability 1, these linear combinations are almost surely linearly independent. Hence, if R1 knows the values of these linear combinations, it can decode data symbols transmitted at time t.

Moreover, by combining the results of previous two lemmas, we observe that R1 can determine the linear combinations $\left\{ \mathrm{LC}_{1i}^{[1]}\left(b - 1, \bar{t}(t) \right) \right\}_{i=1}^{N_1}$ at the end of Block $(b - 1)$, while it can obtain linear combinations $\left\{ \mathrm{LC}_{2i}^{[1]}\left(b - 1, \bar{t}(t) \right) \right\}_{i=1}^{n_{\mathrm{req}}\left(\bar{t}(t) \right)}$ at the end of Block b. Thus, at the end of Block b, R1 can decode all DSs sent to it at time t, which belongs to Phase One of Block $b - 1$. Hence, at the end of Block b, R1 can decode all DSs sent to it over Block $b - 1$. $\blacksquare$

Thus, we conclude that by coding over $T^{\star}$ time slots, $(d_1^{\star}, d_2^{\star})$ DoF can be achieved as desired. This completes the description of our generic retrospective interference alignment scheme.

We will now use this scheme to prove that $P_{o2,3} \in \mathbf{D}^{\mathrm{S}}$ under Case A and $P_{o2,1}, P_{1,3} \in \mathbf{D}^{\mathrm{S}}$ under Case B in the following three subsections.

A. Proof of $P_{o2,3} \in \mathbf{D}^{\mathrm{S}}$ under Case A

Recall that under Case A, Condition 1 holds and

$$M_1' \geq N_2 \frac{N_1 - M_2}{N_2 - M_2},$$

and $P_{o2,3} \equiv (d_1, d_2) = (N_1 - M_2, M_2)$.

We use the generic retrospective interference alignment scheme with the parameters chosen as follows:

$$T = N_2, \ t_1 = N_2 - M_2, \ t_2 = M_2, \ \text{and} \ m_2(\bar{t}) = M_2 \ \forall \bar{t} \in [1 : T].$$

It is easy to verify that this choice satisfies Design Criteria 1 and 2.

In order to choose $m_1(\bar{t})$, $\bar{t} \in [1 : t_1]$, consider the following:

$$\left\lceil N_2 \frac{N_1 - M_2}{N_2 - M_2} \right\rceil \geq N_2 \frac{N_1 - M_2}{N_2 - M_2} \geq \left\lfloor N_2 \frac{N_1 - M_2}{N_2 - M_2} \right\rfloor$$

$$\Rightarrow t_1 \cdot \left\lceil N_2 \frac{N_1 - M_2}{N_2 - M_2} \right\rceil \geq t_1 \cdot N_2 \frac{N_1 - M_2}{N_2 - M_2} \geq t_1 \cdot \left\lfloor N_2 \frac{N_1 - M_2}{N_2 - M_2} \right\rfloor$$

$$\Rightarrow t_1 \cdot \left\lceil N_2 \frac{N_1 - M_2}{N_2 - M_2} \right\rceil \geq N_2(N_1 - M_2) = Td_1 \geq t_1 \cdot \left\lfloor N_2 \frac{N_1 - M_2}{N_2 - M_2} \right\rfloor,$$

This suggests that we can choose $m_1(\bar{t})$, $\bar{t} \in [1 : t_1]$ as

$$m_1(\bar{t}) \in \left\{ \left\lfloor N_2 \frac{N_1 - M_2}{N_2 - M_2} \right\rfloor, \left\lceil N_2 \frac{N_1 - M_2}{N_2 - M_2} \right\rceil \right\}, \ \forall \bar{t} \in [1 : t_1], \text{ such that}$$

$$Td_1 = N_2(N_1 - M_2) = \sum_{\bar{t}=1}^{t_1} m_1(\bar{t}),$$

where $\lfloor x \rfloor$ denotes the largest integer that is less than or equal to x. It can be verified that this choice of parameters satisfies Design Criterion 3. Furthermore, Design Criterion 4 holds since $(N_2 - M_2)t_2 = (N_2 - M_2)M_2$ and $\sum_{\bar{t}=1}^{t_1} m_2(\bar{t}) = M_2 t_1 = M_2(N_2 - M_2)$. Next we prove that Design Criteria 5 is satisfied as well.

First we show that $m_1(\bar{t}) \geq N_1 \ \forall \ \bar{t} \in [1 : t_1]$ for which it is sufficient to prove that

$$\left\lfloor N_2 \frac{N_1 - M_2}{N_2 - M_2} \right\rfloor \geq N_1.$$

This inequality is proved as follows:

$$N_1 \geq N_2 \Rightarrow N_1 - M_2 \geq N_2 - M_2 \Rightarrow \frac{M_2}{N_2 - M_2} + 1 \geq \frac{M_2}{N_1 - M_2} + 1$$

$$\Rightarrow \frac{N_2}{N_2 - M_2} \geq \frac{N_1}{N_1 - M_2} \Rightarrow N_2 \frac{N_1 - M_2}{N_2 - M_2} \geq N_1 \Rightarrow \left\lfloor N_2 \frac{N_1 - M_2}{N_2 - M_2} \right\rfloor \geq N_1,$$

as desired. Now, making use of the fact that $m_1(\bar{t}) - N_1 \geq 0 \ \forall \ \bar{t} \in [1 : t_1]$, we obtain the following:

$$\begin{aligned}
\left| \mathcal{S}_{\mathrm{LC}}(b) \right| &= \sum_{\bar{t}=1}^{t_1} n_{\mathrm{req}}(\bar{t}) = \sum_{\bar{t}=1}^{t_1} \left\{ m_1(\bar{t}) - N_1 \right\} \\
&= \sum_{\bar{t}=1}^{t_1} \left\{ m_1(\bar{t}) \right\} - N_1 t_1 = Td_1 - N_1 t_1 \quad \cdots \text{ by Design Criteria 3} \\
&= N_2(N_1 - M_2) - N_1(N_2 - M_2) \quad \cdots \text{ by the choice of } t_1 \text{ and } T \\
&= M_2(N_1 - N_2) = (N_1 - N_2)t_2,
\end{aligned}$$

which proves that Design Criteria 5 is satisfied. Since all Design Criteria 1-5 hold, we know that $P_{o2,3} \in \mathbf{D}^{\mathrm{S}}$.

B. Proof of $P_{o2,1} \in \mathbf{D}^{\mathrm{S}}$ under Case B

Recall that under Case A, Condition 1 holds and

$$M_1' < N_2 \frac{N_1 - M_2}{N_2 - M_2}.$$

Here,

$$P_{o2,1} \equiv (d_1, d_2) = \left(M_1' \frac{N_2 - M_2}{N_2}, M_2 \right).$$

We use the generic retrospective interference alignment scheme with the following choice of parameters:

$$\begin{aligned}
T &= N_2, \ t_1 = N_2 - M_2, \ t_2 = M_2, \\
m_1(\bar{t}) &= M_1' \ \forall \ \bar{t} \in [1 : t_1], \\
m_2(\bar{t}) &= M_2 \ \forall \ \bar{t} \in [1 : T].
\end{aligned}$$

It is easy to verify that Design Criteria 1-3 are satisfied. Design Criterion 4 holds because $\sum_{\bar{t}=1}^{t_1} m_2(\bar{t}) = M_2 t_1 = M_2(N_2 - M_2) = (N_2 - M_2)t_2$. Consider now the proof that the last criterion is satisfied.

Here, $n_{\mathrm{req}}(\bar{t}) = M_1' - N_1 \ \forall \ \bar{t} \in [1 : t_1]$. Hence,

$$\left| \mathcal{S}_{\mathrm{LC}}(b) \right| = (M_1' - N_1)t_1 = (M_1' - N_1)(N_2 - M_2).$$

Design Criterion 5 because

$$\begin{aligned}
M_1' &\leq N_2 \frac{N_1 - M_2}{N_2 - M_2} \quad \cdots \text{ by definition of Case B} \\
&\Rightarrow M_1'(N_2 - M_2) \leq N_2(N_1 - M_2) = (N_1 - N_2)M_2 + N_1(N_2 - M_2) \\
&\Rightarrow (M_1' - N_1)(N_2 - M_2) \leq (N_1 - N_2)M_2 \\
&\Rightarrow |\mathcal{S}_{\mathrm{LC}}(b)| \leq (N_1 - N_2)t_2.
\end{aligned}$$

Since all Design Criteria hold, $P_{o2,1} \in \mathbf{D}^{\mathrm{S}}$ under Case B.

C. Proof of $P_{1,3} \in \mathbf{D}^{\mathrm{S}}$ under Case B

Here,

$$P_{1,3} \equiv (d_1, d_2) = \left(M_1' \frac{N_1 - N_2}{M_1' - N_2}, N_2 \frac{M_1' - N_1}{M_1' - N_2} \right).$$

Set

$$\begin{aligned}
T = M_1' - N_2, \ t_1 = N_1 - N_2, \ t_2 &= M_1' - N_1 \\
\text{and } m_1(\bar{t}) = M_1 \ \forall \ \bar{t} &\in [1 : t_1].
\end{aligned}$$

In order to choose $m_2(\bar{t})$, consider the following argument.

$$\begin{aligned}
M_1' \leq N_2 \frac{N_1 - M_2}{N_2 - M_2} &\Rightarrow M_1'(N_2 - M_2) \leq N_2(N_1 - M_2) \\
&\Rightarrow N_2(M_1' - N_1) \leq M_2(M_1' - N_2) \\
&\Rightarrow M_2 N_1 - M_2 M_1' + N_2(M_1' - N_1) \leq M_2 N_1 - M_2 N_2 \\
&\Rightarrow M_2(N_1 - N_2) = M_2 t_1 \geq (N_2 - M_2)(M_1' - N_1).
\end{aligned}$$

Therefore, we may select $m_2(\bar{t})$, $\bar{t} \in [1 : t_1]$, such that

$$0 \leq m_2(\bar{t}) \leq M_2 \ \forall \bar{t} \in [1 : t_1] \text{ and } \sum_{\bar{t}=1}^{t_1} m_2(\bar{t}) = (N_2 - M_2)(M_1' - N_1);$$

$$\text{and } m_2(\bar{t}) = M_2, \ \forall \bar{t} \in [t_1 + 1 : t_1 + t_2].$$

It can be easily verified that the above choice of parameters satisfies the first three design criteria. Design Criterion 4 holds because

$$(N_2 - M_2)t_2 = (M_2 - M_2)(M_1' - N_1)\sum_{\bar{t}=1}^{t_1} m_2(\bar{t}).$$

Similarly, it is easy to verify that Design Criterion 5 holds. Therefore, we have that that $P_{1,3} \in \mathbf{D}^{\mathrm{S}}$ under Case B.

V. Conclusion

In this paper, the fast fading MIMO IC is studied under the Shannon feedback setting in which the transmitters are assumed to have perfect knowledge of the channel matrices and the channel outputs, both with a finite delay. Under such a setting, the DoF region is determined with the proof involving in part the demonstration of a key achievability result that in some cases output feedback can improve the DoF region in presence of delayed CSIT. To realize the DoF gains attainable with Shannon feedback, this new achievability scheme not only employs all interference alignment techniques that are feasible with just delayed CSIT, but in addition, also exploits the additional transmitter cooperation that output feedback can induce. This result is further strengthened by identifying scenarios of limited Shannon feedback in which the entire Shannon-feedback DoF region is achievable even under the knowledge of some of the channel matrices and channel outputs at the transmitters. For example, the three DoF regions under just delayed CSIT, just output feedback, and Shannon feedback are proved to be identical for a large class of MIMO ICs. Moreover, while this work obtains the DoF region under output feedback for a large class of MIMO ICs, its complete characterization and its relationship to the delayed CSIT and Shannon feedback DoF regions remains an open problem that merits further investigation.

Appendix A
Proof of Lemma 1

The region $\mathbf{D}^{\mathrm{iCSI\&op}}$ is achievable with just instantaneous CSIT [5]. Thus, it is sufficient to prove that $\mathbf{D}^{\mathrm{iCSI\&op}}$ is an outer-bound to the DoF region of the MIMO IC with instantaneous CSIT and output feedback. Toward this end, recall that the DoF achievable over the point-to-point MIMO channel can not exceed the minimum of the number of transmit and receive antennas [25] (this is referred as the 'single-user' bound). This implies that $d_i \leq \min(M_i, N_i)$ for each $i \in \{1, 2\}$. Consider now the bound on $d_1 + d_2$. Even if both transmitters and both receivers are assumed to cooperate, the total sum-DoF are limited by $M_1 + M_2$ and $N_1 + N_2$ due to the single-user bound. Therefore, $d_1 + d_2 \leq \min\{M_1 + M_2, N_1 + N_2\}$. Now, due to symmetry, it is sufficient to show that $d_1 + d_2 \leq \max(M_1, N_2)$. The proof of this claim, which makes use of techniques developed in [5], is given below.

A. Proof of $d_1 + d_2 \leq \max(M_1, N_2)$ with Instantaneous CSIT and Output Feedback

As stated earlier, it is sufficient to prove that $d_1 + d_2 \leq \max(M_1, N_2)$.

Lemma 13: For the MIMO IC with i.i.d. Rayleigh fading and $N_2 \geq M_1$,

$$d_1 + d_2 \leq N_2$$

when there is instantaneous CSIT and output feedback.

Proof: The proof of this lemma is based on the techniques developed in [5]. This lemma can not however be immediately deduced from [5, Theorem 9] because (i) the model of IC with cooperation studied in [5, Section IV] does not include the case of output feedback considered here and (ii) in the IC considered in [5], the channel matrices are time-invariant and deterministic (not fading) and (iii) the IC is known to be not separable in general [26], [27] . See Appendix A-B for the complete proof. ∎

Thus, as per the above lemma, the required inequality holds when $N_2 \geq M_1$. When $N_2 < M_1$, as argued in [28], [5], we may add $M_1 - N_2$ antennas at R2 (which can not reduce the DoF region), and then apply the above lemma to prove that $d_1 + d_2 \leq M_1$ if $M_1 > N_2$. Therefore, we together have $d_1 + d_2 \leq \max(M_1, N_2)$, as desired.

B. Proof of Lemma 13

This lemma is proved by making use of the techniques developed in [5, Proof of Theorem 9].

Let $U_1(t) \triangleq H_{11}(t)X_1(t) + W_1(t)$ and $U_2(t) \triangleq H_{21}(t)X_1(t) + W_2(t)$. Then we have the following corollary, which is stated using the notation of Section II-A.

Corollary 4 (Lemma 8, [5]): The following is true:

$$\begin{aligned}
\overline{X}_1(n) &\leftarrow \mathcal{M}_1, \mathcal{M}_2, \overline{U}_1(n-1), \overline{U}_2(n-1), \overline{H}(n); \\
\overline{X}_2(n) &\leftarrow \mathcal{M}_2, \overline{U}_1(n-1), \overline{U}_2(n-1), \overline{H}(n); \\
\overline{Y}_1(n), \overline{Y}_2(n) &\leftarrow \mathcal{M}_2, \overline{U}_1(n), \overline{U}_2(n), \overline{H}(n),
\end{aligned}$$

where $a \leftarrow b$ denotes the fact that a is a deterministic function of b.

Proof: Can be proved via induction. ∎

We now apply Fano's inequality assuming that R1 knows the message $\mathcal{M}_2$, and signals $\overline{U}_1(n)$ and $\overline{U}_2(n)$ to obtain

$$\begin{aligned}
nR_1 &\leq I\Big(\mathcal{M}_1; \overline{Y}_1(n), \overline{U}_1(n), \overline{U}_2(n) \Big| \mathcal{M}_2, \overline{H}(n)\Big) + n\epsilon_n \\
&= I\Big(\mathcal{M}_1; \overline{U}_1(n), \overline{U}_2(n) \Big| \mathcal{M}_2, \overline{H}(n)\Big) + n\epsilon_n \\
&= h\Big(\overline{U}_1(n), \overline{U}_2(n) \Big| \mathcal{M}_2, \overline{H}(n)\Big) - h\Big(\overline{U}_1(n), \overline{U}_2(n) \Big| \mathcal{M}_2, \mathcal{M}_1, \overline{H}(n)\Big) + n\epsilon_n,
\end{aligned}$$

where the first equality holds since $\overline{Y}_1(n) \leftarrow \{\mathcal{M}_2, \overline{U}_1(n), \overline{U}_2(n), \overline{H}(n)\}$. Now, following the analysis in [5, Proof of Theorem 9], we get

$$\begin{aligned}
h\Big(\overline{U}_1(n), \overline{U}_2(n) \Big| \mathcal{M}_2, \overline{H}(n)\Big) &\leq h\Big(\overline{Y}_2(n) \Big| \mathcal{M}_2, \overline{H}(n)\Big) + \sum_{t=1}^{n} h\Big(U_1(t) \Big| U_2(t), \overline{H}(n)\Big) \\
h\Big(U_1(t) \Big| U_2(t), \overline{H}(n)\Big) &\leq o(\log_2 P) \\
h\Big(\overline{U}_1(n), \overline{U}_2(n) \Big| \mathcal{M}_2, \mathcal{M}_1, \overline{H}(n)\Big) &\geq n \cdot o(\log_2 P),
\end{aligned}$$

where $o(\log_2 P)$ is constant with n. These bounds give

$$nR_1 \leq h\Big(\overline{Y}_2(n) \Big| \mathcal{M}_2, \overline{H}(n)\Big) + n \cdot o(\log_2 P) + n\epsilon_n. \tag{26}$$

Now Fano's inequality applied at R2 yields

$$\begin{aligned}
nR_2 &\leq I\Big(\mathcal{M}_2; \overline{Y}_2(n) \Big| \overline{H}(n)\Big) + n\epsilon_n \\
&\leq h\Big(\overline{Y}_2(n) \Big| \overline{H}(n)\Big) - h\Big(\overline{Y}_2(n) \Big| \mathcal{M}_2, \overline{H}(n)\Big) + n\epsilon_n \tag{27}
\end{aligned}$$

The desired bound can now be derived by adding inequalities in (26) and (27), and subsequently applying the single-user bound (cf. [5, Proof of Theorem 9]).

APPENDIX B
PROOFS OF COROLLARIES 1 AND 2

A. Proof of Corollary 1

First consider the case of limited Shannon feedback. Since $\mathbf{D}^{lS1}, \mathbf{D}^{lS2} \subseteq \mathbf{D}^{S}$, it is sufficient to prove that the DoF regions $\mathbf{D}^{lS1}$ and $\mathbf{D}^{lS2}$ are achievable when there is limited Shannon feedback of Types 1 and 2, respectively. In other words, we need to prove that the region $\mathbf{D}^{S}_{outer}$ is achievable under two types of limited Shannon feedback, which we do next. Again, assume without loss of generality that $N_1 \geq N_2$.

If Condition 1 does not hold, then $\mathbf{D}^{S}_{outer} = \mathbf{D}^{dCSI}$. Then, as pointed out in [20], the region $\mathbf{D}^{dCSI}$ is achievable when for each $i \in \{1, 2\}$, the ith transmitter knows $H_{ji}(t)$ and $H_{jj}(t)$, $j \in \{1, 2\}$ with $j \neq i$, with a delay. Hence, when Condition 1 does hold, the region $\mathbf{D}^{S}_{outer} = \mathbf{D}^{dCSI}$ is achievable with limited Shannon feedback of Types 1 and 2.

When Condition 1 holds, the region $\mathbf{D}_{\mathrm{outer}}^{\mathrm{S}}$ is shown to be achievable with Shannon feedback of both types by developing a coding scheme in Section IV. It can be verified that this coding scheme works even under limited Shannon feedback of both types.

Hence, together, we have proved that $\mathbf{D}_{\mathrm{outer}}^{\mathrm{S}}$ is achievable with limited Shannon feedback.

For the setting of designable Shannon feedback: Clearly, $\mathbf{D}^{\mathrm{S}} \subseteq \mathbf{D}^{d\mathrm{S}}$. Hence, it is sufficient to prove that the region $\mathbf{D}_{\mathrm{outer}}^{\mathrm{S}}$ is an outer-bound, even under designable Shannon feedback. Toward this end, the proof of Theorem 1 can be easily modified to apply to the general case of designable Shannon feedback.

B. Proof of Corollary 2

The first part of the corollary follows trivially since the setting of Shannon feedback is stronger than that of output feedback. To prove the second part of the corollary, we assume below without loss of generality that $N_1 \geq N_2$; and prove that the region $\mathbf{D}^{\mathrm{S}} = \mathbf{D}_{\mathrm{outer}}^{\mathrm{S}}$ is achievable with output feedback, whenever the inequality $\min(M_1, N_1) > N_2 > M_2$ does not hold (note, with $N_1 \geq N_2$, the second inequality in the statement of the corollary can never be true). This is the goal of the remainder of this sub-section.

Here, by no side-information at the transmitters, we mean the setting where the transmitters have no knowledge whatsoever of the channel states and the channel outputs; and denote the corresponding DoF region by $\mathbf{D}^{\mathrm{no}}$ which is known from the literature [9]–[11], [13].

Throughput the remainder of this subsection, we assume that the inequality $\min(M_1, N_1) > N_2 > M_2$ does not hold. Then

$$\mathbf{D}^{\mathrm{no}} \subseteq \mathbf{D}^{\mathrm{op}} \subseteq \mathbf{D}^{\mathrm{S}} = \mathbf{D}^{\mathrm{dCSI}} = \mathbf{D}_{\mathrm{outer}}^{\mathrm{S}} \subseteq \mathbf{D}^{\mathrm{iCSI}} = \mathbf{D}^{\mathrm{iCSI\&op}}.$$

From [20, Table I], we observe that $\mathbf{D}^{\mathrm{no}} = \mathbf{D}^{\mathrm{dCSI}}$, if the following two inequalities do not hold:

- $M_1 > \max(N_1, N_2)$, $M_2 \geq N_1$, and $M_2 > N_2$ (Case A.I.3 in [20, Table I]); and
- $M_1 > \max(N_1, N_2)$ and $N_1 > M_2 \geq N_2$ (Case A.II.2 in [20, Table I]).

Hence, whenever Cases A.I.3 and A.II.2 do not hold (recall the inequality $\min(M_1, N_1) > N_2 > M_2$ is not true), then

$$\mathbf{D}^{\mathrm{no}} = \mathbf{D}^{\mathrm{dCSI}} \Rightarrow \mathbf{D}^{\mathrm{no}} = \mathbf{D}^{\mathrm{dCSI}} = \mathbf{D}^{\mathrm{op}} = \mathbf{D}^{\mathrm{S}};$$

and thus the corollary holds.

We now proceed to Cases A.I.3 and A.II.2, under which we want to show that $\mathbf{D}^{\mathrm{dCSI}} = \mathbf{D}^{\mathrm{op}}$. Toward this end, for Case A.I.3 and A.II.2, the region $\mathbf{D}^{\mathrm{dCSI}}$ has been shown to be achievable under delayed CSIT by developing two IA-based coding scheme in [20, Section VI] and [20, Section VII], respectively. Although these schemes have been developed there for delayed-CSIT case, they work even with output feedback. Therefore, using the schemes of [20, Section VI] and [20, Section VII], we conclude that $\mathbf{D}^{\mathrm{dCSI}} = \mathbf{D}^{\mathrm{op}}$ under Cases A.I.3 and A.II.2.

REFERENCES

[1] T. Cover and J. Thomas, *Elements of Inform. Theory*. John Wiley and Sons, Inc., 1991.

[2] N. T. Gaarder and J. K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Trans. Inform. Th.*, Jan. 1975.

[3] L. H. Ozarow, "The capacity of the white gaussian multiple access channel with feedback," *IEEE Trans. Inform. Th.*, July 1984.

[4] A. E. Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inform. Th.*, vol. 24, no. 3, pp. 379–381, Apr. 1978.

[5] C. Huang and S. A. Jafar, "Degrees of freedom of the MIMO interference channel with cooperation and cognition," *IEEE Trans. Inform. Th.*, vol. 55, no. 9, pp. 4211–4220, Sep. 2009.

[6] C. Suh and D. Tse, "Feedback capacity of the Gaussian interference channel to within 1.7075 bits: the symmetric case," Jan. 2009, Available: http://arxiv.org/abs/0901.3580.

[7] S. A. Jafar and A. J. Goldsmith, "Isotropic fading vector broadcast channels: The scalar upper bound and loss in degrees of freedom," *IEEE Trans. Inform. Th.*, vol. 51, no. 3, pp. 848–857, Mar. 2005.

[8] A. Lapidoth, S. Shamai, and M. Wigger, "On the capacity of a MIMO fading broadcast channel with imperfect transmitter side-information," in *43rd annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, Sep. 2005.

[9] C. Huang, S. A. Jafar, S. Shamai, and S. Vishwanath, "On degrees of freedom region of MIMO networks without CSIT," Sep. 2009, Available Online: http://arxiv.org/pdf/0909.4017.

[10] C. S. Vaze and M. K. Varanasi, "The degrees of freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT," Sep. 2009, Available Online: http://arxiv.org/abs/0909.5424.

[11] Y. Zhu and D. Guo, "The degrees of freedom of MIMO interference channels without state information at transmitters," Aug. 2010, [Online.] Available: http://arxiv.org/abs/1008.5196.

[12] C. S. Vaze and M. K. Varanasi, "The degrees of freedom region of the MIMO cognitive interference channel with no CSIT," in *IEEE Intnl. Symp. Inform. Th.*, Austin, USA, June 2010.

[13] ——, "A new outer-bound via interference localization and the degrees of freedom regions of MIMO interference networks with no CSIT," *submitted, IEEE. Trans. Inform. Th.*, May 2011, Available: http://arxiv.org/abs/1105.6033.

[14] C. S. Vaze, S. Karmakar, and M. K. Varanasi, "On the generalized degrees of freedom of the MIMO interference channel with no CSIT," in *Proc. Intnl. Symp. Inform. Th.*, St. Petersburg, Russia, Aug. 2011.

[15] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," Oct. 2010, Available: http://arxiv.org/abs/1010.1499.

[16] C. S. Vaze and M. K. Varanasi, "Degrees of freedom region for the two-user MIMO broadcast channel with delayed CSI," in *Proc. IEEE Intnl. Symp. Inform. Th.*, St. Petersburg, Russia, Aug. 2011, Available: http://arxiv.org/abs/1101.0306.

[17] M. J. Abdoli, A. Ghasemi, and A. K. Khandani, "On the degrees of freedom of three-user MIMO broadcast channel with delayed CSIT," in *IEEE Intnl. Symp. Inform. Th.*, St. Petersburg, Russia, Aug. 2011.

[18] H. Maleki, S. A. Jafar, and S. Shamai, "Retrospective interference alignment," Sep. 2010, Available: http://arxiv.org/PS_cache/arxiv/pdf/1009/1009.3593v1.pdf.

[19] C. S. Vaze and M. K. Varanasi, "The degrees of freedom of the 2x2x2 interference network with delayed CSIT and with limited shannon feedback," in *Proc. 49th Annual Allerton Conference on Communications, Control and Computing*, Monticello, IL, USA, Sep. 2011, submitted for review, July, 2011 (see arXiv:1109.5790)",.

[20] ——, "The degrees of freedom region and interference alignment for the MIMO interference channel with delayed CSI," *submitted, IEEE Trans. Inform. Th.*, Jan. 2011, Available: http://arxiv.org/abs/1101.5809.

[21] H. L. Royden, *Real Analysis*. Prentice Hall, 1988.

[22] C. S. Vaze and M. K. Varanasi, "Can output feedback improve the degrees of freedom of a wireless network with delayed CSIT?" in *Proc. 49th Annual Allerton Conference on Communications, Control and Computing*, Monticello, IL, Sep. 2011, submitted for review, July, 2011 (see arXiv:1109.5779).

[23] R. Tandon, S. Mohajer, H. V. Poor, and S. Shamai, "On interference networks with feedback and delayed CSI," Sep. 2011, Available: http://arxiv.org/abs/1109.5373.

[24] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge Univ. Press, 1985.

[25] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *Euro. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[26] V. R. Cadambe and S. A. Jafar, "Parallel gaussian interference channels are not always separable," 2008, Available: http://newport.eecs.uci.edu/s̃yed/papers/int_freqsel.pdf.

[27] L. Sankar, X. Shang, E. Erkip, and H. V. Poor, "Ergodic fading interference channels: Sum-capacity and separability," 2009, [Online.] Available: http://eeweb.poly.edu/ēlza/Publications/it09c.pdf.

[28] S. A. Jafar and M. J. Fakhereddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. Inform. Th.*, vol. 53, no. 7, pp. 2637–2642, Jul. 2007.